

Defeated but not deterred

Animal-rights protesters have helped to end plans for a primate research centre at the University of Cambridge. But despite the activists' triumphant soundbites, their victory is unlikely to be repeated elsewhere.

It doesn't make pleasant reading, but the gloating should come as no surprise. When the University of Cambridge decided last week to scrap plans for a primate research centre, animal-rights protesters who campaigned against the facility were in bellicose mood. "An important message has been sent out to universities across the UK," activists' group Stop Primate Experiments at Cambridge said in an e-mail to supporters. "If the mighty Cambridge can only put up a pathetic fight, then they have no chance if we were to turn our attention to them."

For any university that conducts animal research — not just the small number that use primates — this sounds like a worrying challenge. Activists clearly helped to derail the Cambridge centre. Fears about protests prolonged the planning inquiry. The need for security at the site bumped up running costs. Concerns that building workers would be threatened led some to wonder whether it could be built. When costs jumped by £8 million (US\$15 million) to £32 million, the university had to pull the plug.

But that does not mean that the protesters are guaranteed other victories. Primate research is already being done in Britain, and does not need to be moved to centralized facilities. The activists also ignore public and media opinion towards animal experiments, which is more positive than they dare acknowledge. Universities and funding bodies should not prepare for war with animal-rights activists, but continue with the lobbying that is winning people round.

There are several reasons why the Cambridge débâcle is unlikely to be repeated. Funding bodies and universities are considering different approaches after observing the planning process there. Researchers who argue against the need for centralized animal labs, such as neuroscientist Colin Blakemore, who has faced death threats for advocating animal research, may now hold sway.

Blakemore believes that embedding animal facilities in research groups makes for better science and better public relations. Animal work complements the other techniques of the life sciences, from *in vitro* studies to computer modelling. Separating it from mainstream science, he says, hampers research and makes animal research appear backwards-looking. As chief executive of the Medical Research Council (MRC), Blakemore, together with the Wellcome Trust, the major funder for the Cambridge centre, is exploring how Cambridge can continue to do primate work. From the MRC's point of view, this means funding the groups that already have primate facilities.

Some universities will nevertheless decide that centralized animal facilities are the best way to improve animal care. They are likely to face threats of violence, but there is every reason to think they can resist. Opinion polls suggest broad support among the public for strictly regulated animal research. And the reaction of the media towards the Cambridge decision has been overwhelmingly pro-science. Grilled on *Today*, the BBC's flagship radio news show, one activist appeared badly out of his depth.

Good public relations could head off future protests. The Research Defence Society (RDS), for example, regularly discusses animal experiments with the public and explains why they are needed. This must be a better way forward than one option considered for the Cambridge facility — retreating to a fortified lab at a secure but secretive site such as Porton Down, owned by the Ministry of Defence.

Contrast this with the situation in the Netherlands and Switzerland, where animal researchers sometimes show members of the public round their lab during open days. No university in Britain could risk running such events. But unless we pursue a policy of openness, such events will never be possible, and the Cambridge decision will not be the campaigners' last victory. ■

Sound thinking

For city dwellers, maps of noise pollution are a good example of what science can do to improve quality of life.

A three-dimensional (3D) computer graphic can be worth a thousand words. The European Union's efforts to create 3D noise maps of its cities, railways and airports (see page 480) have given experts the tools both to assess the noise pollution to which its citizens are exposed, and to simulate how it can be muffled by intelligent planning of buildings, acoustic shields and speed limits. And less sophisticated versions of the maps have been made available to the public, bringing science closer to the people.

The maps are stunning, with average noise levels superimposed on 3D visualizations of entire cities in coloured contours ranging from pale-green — less than 45 decibels — to deep blue for greater than 79 decibels. They are a powerful way for politicians and the public to better understand noise, and to participate more meaningfully in urban-development and noise-action schemes that were previously limited to experts. Parisians, for example, can now go to the web to look up the noise levels in their street, or even on the façade of their apartment. Where anecdotal evidence and parochial

complaints once reigned, the public can now both better argue its own particular gripes and be convinced at a glance of the broader benefits of unpopular controls, such as stricter traffic speed limits.

There is good reason to make a din. Some 80 million Europeans already suffer noise above 65 decibels, enough to cause lost sleep, stress, high blood pressure and even heart attacks, and a further 170 million endure levels that are simply annoying. And traffic of all kinds — the principal source of noise pollution — is set to increase.

Noise modelling is not rocket science, but more research is needed to improve the quality of the geographical information systems and noise data used and to refine the maths and physics of the models. Materials science could bring us quieter road surfaces and more effective acoustic shields, and engineering and fuel science could give us quieter engines. The effect of noise on people and their health also needs to be better understood. The European Union should be congratulated for setting about drafting such a research agenda; other countries should sit up and listen. ■





Drug control
Experts seek review
of antidepressant use
in children
p474



Flat response
Cold comfort for
science in US budget
for 2005
p475



Going global
German universities
jockey for position
at the cutting edge
p477



Gutted
Exploding sperm
whale halts the
traffic in Taiwan
p478

Fear of human pandemic grows as bird flu sweeps through Asia

Alison Abbott and Helen Pearson

Public-health officials and lab researchers are this week engaged in a frantic battle to contain the escalating outbreak of bird flu in Asia, and to minimize the risk of it developing into a human flu pandemic.

The first cases of apparent human-to-human transmission of the avian flu virus were reported in Vietnam on 1 February, heightening public-health concerns. As officials struggle to contain the avian outbreak by orchestrating the massacre of chicken flocks, researchers around the world are preparing the way for the mass production of a vaccine that could help to contain a human pandemic.

Individual human-to-human transmission of the virus doesn't necessarily mean that it has evolved into a strain that will transmit readily enough to cause a significant outbreak, never mind a pandemic. But should a pandemic occur, many experts say that current public-health plans are unlikely to prevent large-scale loss of life. Previous human flu pandemics, such as the 1918 outbreak that killed about 40 million people, originated in birds.

The present outbreak of the highly pathogenic H5N1 virus — named for the



Cutting the links: health officials are culling chickens in a concerted effort to halt the spread of bird flu to humans.

subtypes of its key surface proteins, haemagglutinin and neuraminidase — has ripped through flocks of domestic poultry in south-east Asia and spread to at least 14 people, killing 11 of them.

In efforts to curb the spread of the virus, the eight countries affected so far (see map, opposite) have slaughtered 20 million chickens. Experts say that culling infected birds is the only way to extinguish the virus before it infects more people. This can reduce the risk that avian and normal human flu meet and exchange genes in an individual, creating a strain with the ability to spread

easily between people (see 'Mixed messages', opposite). Officials don't put much stock in vaccinating the region's vast chicken flocks, especially as this would not eliminate infectivity — healthy birds that are still shedding the virus would undermine the effort.

The World Health Organization (WHO) is working on the first stage of a five-point action plan for dealing with pandemic influenza, says Klaus Stöhr, project leader of the organization's global influenza programme. This involves seeking to contain the avian outbreak while monitoring human-to-human transmission.

The US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, a laboratory at St Jude Children's Research Hospital in Memphis, Tennessee, and the National Institute for Biological Standards and Control in London are working with the WHO to prepare a human vaccine. They expect a prototype H5N1 vaccine strain to be engineered shortly and safety-tested in ferrets within two months. This will then be handed to manufacturers for clinical testing, which should take up to another two months. If the virus starts spreading between people, mass production would begin, which would take an extra four to six months.

Lack of infrastructure hampers virus monitoring

In the capital of Cambodia, Phnom Penh, only six adults and children have been tested for influenza this winter. "Routine surveillance is just not possible," says Sean Tobin, a World Health Organization epidemiologist stationed there.

That's the reality facing public-health officials in the poorer countries where avian flu may or may not mutate into a human pandemic. And for all their talk of control and mitigation strategies, health officials acknowledge that there's little that can be done on the ground in Cambodia, Laos or other countries that lack even a rudimentary public healthcare system.

According to a spokesman for the World Organisation for Animal Health, Laos and Cambodia lack many of the essentials needed

to control an outbreak, including sufficient numbers of veterinarians, transport to get samples to the laboratory — or indeed laboratories themselves. He says that infected populations of chickens will continue to go undetected in these countries.

Rapid diagnostic tests to identify avian flu in chickens are available at national centres in Vietnam and Thailand, but not in their poorer neighbours. In Cambodia, it took almost two weeks from the detection of suspected cases of avian virus in chickens — with subsequent tests done by the Pasteur Institute in Paris — for the government to confirm the report. It took officials in Laos eight days after suspected cases were detected before they could announce results

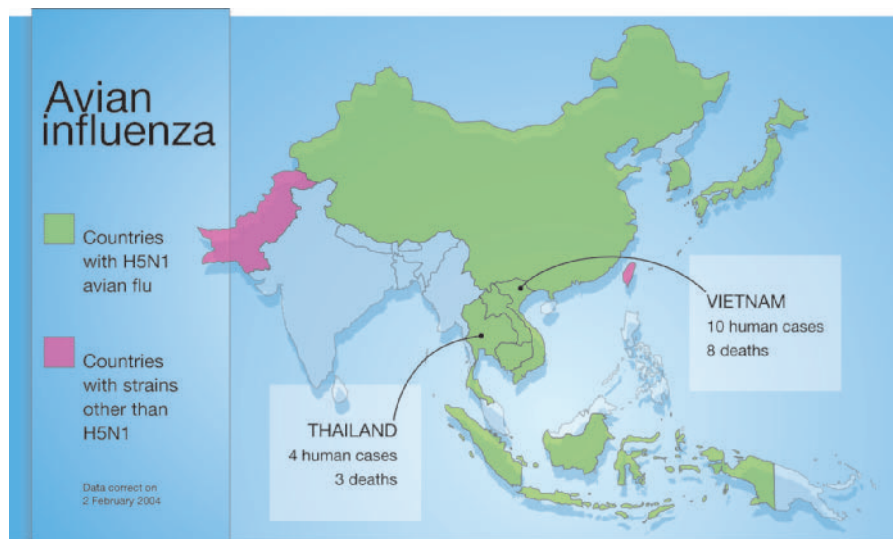
from laboratories in Thailand and Vietnam.

Cambodia and Laos are also unable to test for human influenza subtypes. These tests are needed to isolate patients infected with avian virus from those with the normal flu.

Last year's SARS epidemic was a learning experience for Cambodia, says Tobin. "Some health officials in the provinces are used to working with case definitions that we provide," he says. "But getting information down to the provincial level is notoriously difficult in this country." And whereas the main SARS risk was restricted to areas visited by foreigners, the risk of avian flu is nationwide, affecting regions where public health resources are at their most meagre.

David Cyranoski

D. LONGSTREATH/AP



The lengthy production schedule means that in the worst-case scenario, the first wave of a pandemic could almost be over before a vaccine is ready, some experts warn. "Most of the damage will have been done by the time a vaccine is widely available," says Rino Rappuoli, head of research at Chiron Vaccines in Siena, Italy, one of the major companies in Europe working on pandemic flu vaccines. Even so, a vaccine might be useful in fighting any subsequent 'waves' of the pandemic.

Stöhr concedes that "vaccines will not be ready for those at the epicentre of any pandemic". Antiviral drugs will be the first line of medical intervention, he says, but these may also be in short supply. Preliminary tests done by the CDC have shown that the H5N1 virus is resistant to a traditional group of drugs, including amantadine and rimantadine, that block virus replication.

Short supply

A newer and more expensive class of drug that cripples the neuraminidase enzyme in flu viruses is effective against H5N1 strains. The most common version of these drugs is oseltamivir (Tamiflu), made by Swiss-based drug company Roche. But the company won't say how many doses of Tamiflu it has or reveal its capacity to manufacture more, says spokesman Terence Hurley. He claims that Roche is cooperating with the WHO and offering Tamiflu at a reduced price for emergency use in poor countries.

Other experts doubt whether stocks of Tamiflu would be sufficient in the face of a pandemic. "It's in very limited supply and would quickly run out," says epidemiologist Ira Longini at Emory University in Atlanta, Georgia, who has modelled Tamiflu's potential impact. Prompted by the severe flu season this winter, the CDC is already building up a significant stockpile of Tamiflu. Japan and some European countries are also negotiating with Roche to get stocks of the drug.

Meanwhile, efforts in southeast Asia are aimed at lowering the rate of transmission to

humans by culling chickens as quickly and safely as possible. Stöhr is concerned that many of those doing the culling don't have the recommended protective clothing, and aren't being vaccinated against the circulating strain of normal human flu, or taking prophylactic antiviral drugs. These steps would reduce the chances of the avian flu combining with human flu strains in their bodies. "We need to get rid of the natural reservoir of H5N1, but we need to do it safely," says Stöhr.

Some researchers hope that the current outbreak will add impetus to their calls for governments to be better prepared. They would like to see improved surveillance of circulating strains of animal flu, large-scale manufacture of vaccines to possible pandemic strains before human-to-human transmission is seen, or stockpiling of antiviral drugs.

Many countries are drafting such schemes. The German health authorities, for example, are finalizing a plan that would routinely prepare prototype vaccine strains based on subtypes of the haemagglutinin surface proteins circulating in animals or birds. "This could save months in vaccine preparation if a pandemic threat arises in the future," says Reinhard Kurth, president of the Robert Koch Institute in Berlin.

The Canadian government has already implemented part of its pandemic influenza plan: it has a contract with Quebec-based Shire Biologics that requires the company to be ready to manufacture a pandemic vaccine. The company could produce 6 million doses per month once it is supplied with the prototype vaccine strain, says Health Canada spokesperson Emmanuel Chabot in Ottawa, Ontario.

Other countries should invest in similar schemes to ready themselves for future flu pandemics, experts say. "We are going to face this over and over again," predicts public-health researcher Scott Layne of the University of California, Los Angeles.

For more coverage see:

► www.nature.com/nature/focus/birdflu

Mixed messages

Public-health officials badly need to know whether a strain of flu virus is likely to emerge that can spread readily from people to people — but researchers can't tell them. This knowledge gap is stirring debate about whether animals should be co-infected with bird and human flu viruses in the laboratory in an attempt to get a better grip on the biological mechanisms involved.

"It would give us some idea of how concerned we should be," says Richard Webby, a virologist at St Jude Children's Research Hospital in Memphis, Tennessee. "But it's a question of whether these experiments should be done — you would potentially be creating a pandemic virus."

A strain of avian flu that could be transmitted to humans might emerge if the avian H5N1 virus and an existing human flu strain both infect the same cell in the same person or bird. Mixing of genes from the two strains could then result in a human-transmissible virus. "It's a random event that we can't just predict, given what we currently know," says David Heymann, an infectious-disease expert at the World Health Organization in Geneva.

Efforts to model the scenario on a computer are hampered by a lack of information on how many birds and people harbour each virus. Additionally, says virologist Robert Webster, also at St Jude, "We know zippo about what genes are involved in transmissibility."

Researchers are worried that it is only a matter of time before the two viruses meet and mix. And some think that the number of people infected with avian flu in southeast Asia is greatly underestimated by official statistics. "The more exposure to avian viruses, the more chance you would have that such an event would occur," says virologist Albert Osterhaus of Erasmus University in Rotterdam, the Netherlands.

Human and avian influenza viruses are already being mixed in labs that research flu vaccines — but only using non-pathogenic strains of human virus and less virulent forms of the bird virus. To model the pandemic threat, researchers need to know the conditions under which viruses mix their genes to generate the most deadly combinations.

"It is possible to create a virus with pandemic potential," says Osterhaus. But such an experiment "would have to be done in an extremely high-security laboratory", says Heymann. "You could create a monster."

And some are sceptical of what would be gained from better laboratory data. Real pandemics are influenced by stochastic events that can't be predicted, points out Roy Anderson, an epidemiologist at Imperial College, London, who is trying to model the probability of such combinations. "It's exceedingly difficult and I'm not optimistic that we'll produce something of value," he says.

Carina Dennis

Drug suicide risks prompt call for FDA action

Erika Check, Washington

A scientific advisory panel has joined a chorus of patient groups and consumer advocates this week in calling for the US Food and Drug Administration (FDA) to curb the use of certain antidepressant medications in young people.

"We would like the FDA to go ahead and issue a stronger warning to clinicians regarding the possible risks of these medicines," says psychiatrist Matthew Rudorfer of the National Institute of Mental Health in Bethesda, Maryland. Rudorfer chairs a committee that met on 2 February to advise the FDA on certain antidepressant medications in patients aged 17 and under. But the committee said that the drugs should not be banned.

The meeting followed a decision by British regulators on 10 December last year to ban doctors from treating children with five antidepressant drugs classified as 'selective serotonin reuptake inhibitors', or SSRIs. The British decision was based on an assessment of drug-company data suggesting that the drugs were not proven to be effective in young people, but marginally increased the risk that they would commit suicide.

The UK Medicines and Healthcare Products Regulatory Agency ruled that only one drug of this type — fluoxetine, better known as Prozac — was beneficial enough to justify the risk of side effects, and should be pre-



Terri Williams (left) waits to give testimony at a public hearing; her 14-year-old son Jacob committed suicide while on antidepressants.

scribed to children. In the United States, Prozac is the only SSRI that has been approved by the FDA for use in children, but others are nonetheless frequently prescribed by physicians.

The committee meeting heard emotional testimony from parents whose children had committed suicide while on SSRIs. Some parents raised concerns about the accuracy of published literature on the drugs.

Thomas Laughren, an FDA official responsible for psychiatric drugs, told the committee that the FDA's approach to the drugs was based on identical data to those considered by British regulators. He says

that there's "enough of a signal here" to suggest that something is amiss.

But the FDA says that the data are inconclusive, and wants more information before following the British ban. It has therefore asked a group of researchers at Columbia University, New York, to help it to reanalyse the data from the drug-company studies.

Scientists on the committee predicted that the outside analysis would not answer the basic question of whether the drugs cause children to try to kill themselves. They pointed out that corporate studies are not designed to measure how the medications affect the risk of suicide. "I don't have a lot of hope for you being able to get good information out of the planned reanalysis," says Judith O'Fallon, a panel member and biostatistician at the

Mayo Clinic in Rochester, Minnesota.

The panel said that the FDA should still go ahead with its reanalysis of the data — but should also do more to caution physicians about the risks posed by the drugs. FDA officials said they would consider this advice, but that the agency had already sent out a warning to doctors last October and was not planning major regulatory action before completing its reanalysis this summer. The officials added that smaller steps could be taken in the meantime, such as changing the labelling of the drugs, and including more information in leaflets that accompany them. ■

Europe urged to move on transgenic crop imports

Quirin Schiermeier, Munich

The president of the European Commission last week called for an end to Europe's effective moratorium on the growth and importation of genetically modified crops.

Supporters of agricultural biotechnology are hoping that a letter from Romano Prodi to the commissioners of the European Union's (EU's) 15 member states will mark the beginning of the end of the moratorium. The letter, sent on 28 January, urges member states to restart the approval process for the crops to avoid a trade war with the United States. But the jury is still out on whether Prodi's message will make much difference.

The EU hasn't given the go-ahead for the importation or cultivation of any genetically modified crop since October 1998, and last August a World Trade Organization panel

was set up to consider a complaint from the United States, Canada and Argentina that the go-slow violates its trading rules.

Last year, the EU passed laws stipulating the labelling and traceability of the crops. The laws come into effect in April. "Now that the necessary legislative decisions have been taken," Prodi wrote in his letter, "it is important to demonstrate to the European public and to our trade partners that the EU system of authorisation is working as designed."

Currently, 22 applications are awaiting approval across the EU. The first in line — and an acid test of the new policy — is a request to import an insect-resistant sweetcorn variety called Bt11, developed by the Swiss firm Syngenta and grown in the United States.

In December, a regulatory committee

failed to reach the qualified majority required to permit importation of the corn. The matter will now go to a council of ministers from member states, which has three months to make a decision. But if the council fails to achieve the necessary majority either in favour or against — as many observers predict — the European Commission will decide. If that happens, imports are likely to start as early as April.

One crop that is unlikely to be approved is herbicide-tolerant oilseed rape, marketed by Bayer CropScience in Monheim am Rhein, Germany, and submitted for approval to the Belgian government. Ministers ruled against the application on 2 February after studying the results of a British report that found that the herbicide-spraying regime associated with oilseed rape could harm the environment. ■



Balancing the books: President Bush shares his budget proposal with members of his Cabinet.

Bush's belt-tightening budget offers science slim pickings

Geoff Brumfiel, Washington

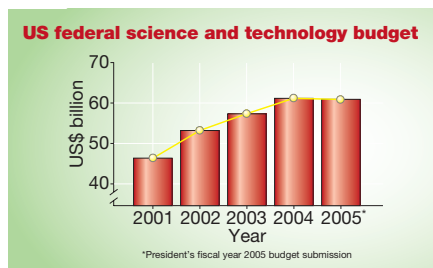
A decade of strong growth in US research funding came to an abrupt halt on 2 February, when President Bush released a budget proposal that attempts to confront the nation's massive financial deficit.

The budget for the 2005 fiscal year, which begins on 1 October, offers little new money for researchers. Funding for the National Institutes of Health (NIH) and the National Science Foundation (NSF) would struggle to keep up with inflation, and programmes at most other major agencies are cut. Supporters of science in Congress have vowed to push for more funding as the budget takes final shape there this summer — but few lobbyists hold out much hope that they will succeed.

Administration officials put a positive spin on the numbers. John Marburger, the president's science adviser, told a briefing at the National Academies in Washington DC that the budget allocates a record \$132 billion to research and development in 2005 — 5% more than last year.

But science advocates point out that almost all of that extra money goes to evaluating military equipment. The budget for 'federal science and technology' — the definition set by the National Academies as the measure for innovative research and development — would be \$60.4 billion, 0.5% less than 2004 (see chart). "This is a slight, across-the-board cut for science," says Mike Lubell, director of public affairs at the American Physical Society.

At the NIH, which funds most biomedical research in the United States, funding would increase by 2.6% to \$28.6 billion. Much of the new money is directed at biodefence research, which would grow by 7.5% to \$1.7 billion. NIH director Elias Zerhouni called the current budget climate "difficult" because



the agency must contend with bioterrorism, while continuing to fund multi-year research grants. The NIH plans to fund only about 250 new research grants in 2005. "They are spreading resources more thinly," says Pat White, head of federal relations at the Association of American Universities.

The NSF, which distributes most non-biomedical research grants at US universities, would see a 3% increase to \$5.75 billion. This is above what some other agencies are being offered, but is far below the level suggested by a law passed in 2002, which recommended doubling the NSF's budget over five years. "The NSF has fared pretty well, relatively speaking," director Rita Colwell says. But in a show of frankness unusual for an agency chief, Colwell adds that "it would be disingenuous to say that this is all that we had hoped for."

The Department of Energy's office of science, which funds most US physics, sees its budget cut by 2% to \$3.4 billion.

With both Congress and the administration trying to tame the half-a-trillion-dollar deficit ahead of November's elections, science advocates will have their work cut out getting more money before the budget is finalized. In the words of one senate staffer: "This is going to be ugly."

Additional reporting by Paul Smaglik, Tony Reichhardt and Erika Check.

Boost for biodefence

The National Institutes of Health's budget includes \$150 million to build 20 'biosafety level 3' secure labs for investigating hazardous biological materials. Once completed, they will conduct about 200 projects finding ways to protect the public from bioterrorism.

At the Centers for Disease Control and Prevention in Atlanta, Georgia, the overall 2005 budget is down by 1%. But a new, \$130-million biosurveillance programme is proposed to monitor the 'quiet' spread of diseases by tracking non-patient data, such as sales of particular pharmaceuticals.

Military might

The Pentagon's overall budget is boosted by 7% to \$400 billion, but researchers and grant recipients won't see any of the extra cash.

Basic and applied research at the defence department — the main source of support for computer science and engineering research in US universities — would be cut by 11% to \$5.2 billion.

But funding for the deployment of missile defence systems — whose efficacy has been questioned by US physicists (see *Nature* 424, 240; 2003) — will rise by 13% to \$10.3 billion.

Moonbound, eventually...

Since President Bush announced his goal to return astronauts to the Moon (see *Nature* 427, 273; 2004), the question has been how NASA will pay for it. A little light was shed on this in the agency's budget proposal, which sees its overall allocation rise by 5.6% to \$16.2 billion.

Most of the money for the new programme will come from converting existing rocket development work to designing new vehicles and phasing out the space shuttle. Scientists will have to make sacrifices too. A proposed 'Beyond Einstein' initiative to study dark matter and black holes would be deferred. Spending on Earth science would drop by 8% and space science would grow at a slower rate than projected last year. But Mars-related science and technology is up by 16% to \$691 million.

...but less joy on Earth

Environmental science programmes are among those hardest hit in the entire proposal. The budget for science and technology at the Environmental Protection Agency (EPA) is cut by 12% to \$577 million. Research at the National Oceanic and Atmospheric Administration is down 3% to \$350 million, and research at the US Geological Survey is down by 2% to \$920 million.

Mark Udall (Democrat, Colorado), senior minority party member on the House subcommittee overseeing environmental research, says: "I am very disappointed. This is a budget lacking vision."

Suicide-inquiry fallout 'could gag' scientists

Jim Giles, London

When the report into the death of UK government scientist David Kelly was published on 28 January, a painful picture emerged of the biowarfare researcher's last days. Kelly, who committed suicide last July after being caught in the crossfire between the government and the BBC, was revealed as an extremely private man who felt unable to discuss the extraordinary stress he was under, even with his family.

But for the people who worked with him, Kelly was a valued expert and a vital source of accurate information. In the aftermath of Lord Hutton's report into his death, many fear that such sources will be hard to find as government scientists become less willing to share what they know. "I used to meet Kelly at conferences where there was a relatively free flow of information," says Malcolm Dando, a specialist in international security at the University of Bradford. "It will be to the detriment of my work if that flow ends."

Kelly was a leading expert on biological weapons. As former head of microbiology at Britain's Chemical and Biological Defence Establishment at Porton Down, Wiltshire, he was credited with reorientating the facility towards monitoring proliferation after Britain renounced chemical and biological weapons in 1972. But Kelly really made his name in the tight-knit international community of specialists in weapons of mass destruction in the early 1990s, when he investigated two weapons programmes.

On a series of visits to Russian biotechnology laboratories, as part of an Anglo-American inspection programme, Kelly helped to uncover a clandestine biological-weapons programme that had continued there long



David Kelly committed suicide in July as a row raged over UK intelligence on Iraqi weapons.

after the Soviet Union publicly renounced such weapons. And after the 1991 Gulf War, his patient questioning of Iraqi scientists as part of a United Nations team helped to force the country's government to admit that it held stocks of biological weapons.

The insights that Kelly gained made him an excellent contact for security researchers. "His presence at seminars helped us understand the issues we study," says Richard Guthrie, an expert on chemical and biological weapons at the Stockholm International Peace Research Institute.

Like others in his field, Guthrie suspects that government scientists all over the world may be less willing to talk after seeing the pressure that Kelly was placed under.

Kelly took his own life after being named as the source for a news story on BBC radio by reporter Andrew Gilligan, who alleged that

the British government had distorted intelligence about Iraq's weapons programmes to make the case for military intervention. Kelly admitted meeting Gilligan but denied he was the source for the story — a denial that was publicly exposed as false in circumstances that probably led to his suicide.

Although Kelly breached civil-service rules by discussing intelligence matters with Gilligan and other journalists, he was officially authorized to discuss technical matters with them. Kelly regularly attended academic seminars, but Guthrie fears that others will now think twice about talking openly at seminars. "If I was a civil servant and someone came up to me during coffee, I might not want to talk to them," he says. "I might not even go to the seminar."

Others caution against reading too much into Kelly's case. Graham Pearson, a retired head of Porton Down, says that very few government scientists have access to the kind of intelligence that Kelly had, so they won't feel under pressure to curtail their interactions with academics.

Kelly's death was nonetheless a heavy blow to those who study biological-weapons proliferation. This small group remembers him as a helpful and open colleague — but the inquiry starkly revealed the kind of double life that he was obliged to lead.

Some of the most poignant testimony came from his widow, who revealed that as a media storm raged around Gilligan's report and its then-unnamed source, Kelly remained silent at home. He only hinted at his role on 8 July — more than a week after he had revealed it to the Ministry of Defence and a full month after Gilligan's broadcast. ■

♦ www.the-hutton-inquiry.org.uk

Fusion meeting shelved as site decision slides

Declan Butler, Paris

Ministers from the partners in ITER, a US\$5-billion international fusion experiment, have postponed a meeting scheduled for this month to select a site for the project.

The delay has been caused by deadlock over the choice between France or Japan as host for the project, which would seek to prove the principle of creating fusion energy by heating plasma constrained by a magnetic field.

Negotiations over ITER's site have been stalled since a ministerial meeting in Washington in December, when the United States and Korea backed the Japanese site at Rokkasho, and China and Russia supported the European Union's site at Cadarache in

France. Negotiators asked France and Japan to answer questions on the technical merits of each site, and to consider making the experimental reactor into part of a broader, international fusion-research package.

The ministers agreed to reconvene in February but, with no agreement in sight, another failed meeting "would have given the impression that the whole ITER process was falling apart", according to an official close to the negotiations. Now ITER's backers hope a delay will help draw the political sting from the negotiations, which have been overshadowed by the perception that the United States is supporting Japan out of a desire to punish France for its stance over last year's invasion of Iraq.

The broader research programme would

look, for example, at how the reactor walls of a future fusion power plant stand up to radiation damage, and at superconducting magnets. Such a programme would cost about \$800 million, and would be seen as a way of compensating the country that doesn't get to house the main experiment.

In the absence of a ministerial meeting, government officials connected to ITER will meet in Vienna on 21 February to try to break the deadlock.

Meanwhile, the European Union and Japan are engaged in talks in a bid to find a way forward. One European official says that the European Union is offering South Korea and Japan support for projects ranging from genomics to neutron science, if they will support the French bid to host ITER. ■



Humboldt University of Berlin: Germany is concerned over the global standing of such institutions.

Universities battle for extra funds in bid to boost quality

Quirin Schiermeier, Munich

Germany's top universities are being asked to compete for top-up grants from the federal government to help them pack more of a punch internationally.

Five winning institutions will each get extra funding of €50 million (US\$60 million) every year for five years from 2006, in a bold scheme designed to help them compete on a global stage with the likes of Harvard University and the University of Tokyo.

A competition for the money will start in the summer, Germany's science minister Edelgard Bulmahn announced at a meeting in Berlin on 26 January. She called on all of Germany's 100 or so research universities to apply by setting out a plan of how they would use the money to improve teaching and research. The winners will be chosen next year by a committee of German and foreign researchers and administrators.

"We urgently need more money for research, but we also need more quality for our money," Bulmahn says.

Other participants at the meeting included Horst Störmer of Columbia University in New York, who won the 1998 Nobel physics prize, and Chancellor Gerhard Schröder, who said last month that Germany must do more to promote its elite universities (see *Nature* **427**, 271; 2004).

The federal government has promised €1.25 billion to the competition. One option for finding the money is to sell part of the German central bank's gold reserves (see *Nature* **427**, 386; 2004).

In 1996, the government launched a similar competition, called BioRegio, in which different regions of the country competed for biotechnology funding for grants and equipment. Policy-makers think that this initiative had some success in strengthening German biotechnology.

Critics argue that it will take more than a cash competition to reform the country's university system. "It is not sufficient that universities send around applications boasting how well they can perform," says Dagmar Schipanski, science minister of the east German state of Thuringia. She thinks it would be better just to give the money to the DFG, the country's main research funding agency, so that all researchers could compete for it.

But Peter Gaehtgens, president of the German Conference of University Rectors, says he is optimistic that the competition will help to increase the international standing of Germany's top universities. In a Chinese survey of research universities released last month, no German university came in the top 40.

"Fresh money is always welcome," says DFG president Ernst-Ludwig Winnacker. "But it takes more than that. German university laws still promote equality, rather than competition for the best students and young scientists."

Winnacker, Gaehtgens and the heads of several other research organizations meet with Bulmahn later this month to hammer out the details of the competition, and to discuss further steps that could be taken to modernize Germany's universities. ■

Extinction meeting kicks off Japan's plans for networking

David Cyranoski, Okazaki

East Asian researchers can sometimes feel cut off from the merry-go-round of small but prestigious meetings that help to further the careers of their colleagues in Europe and North America. Now a series of conferences in Japan aims to give these scientists similar benefits.

On 25–30 January, at the First Okazaki Biology Conference, 70 researchers from 10 countries met in the coastal city of Okazaki on Japan's main island to discuss the biology of species extinction.

The conference's organizers hope that it is the first of what will become a significant series, modelled loosely on the Gordon Research Conferences held in the United States. "There's nothing like the Gordon Conferences in this region," says Yoshitaka Nagahama, a developmental biologist at Okazaki's National Institute for Basic Biology, and chair of the series.

Motoya Katsuki, the institute's director, who thought up the idea of the conferences, says the plan is to bring together "many researchers who are climbing different sides of the same mountain and can't see one another".

The extinction meeting brought together specialists in long-term climate modelling, the geographical distribution of animals, and the origin and evolution of species.

But getting Japanese researchers to join the international networks that serve to build new fields of research won't be easy: much of the meaningful exchange at small meetings goes on informally in conversations between sessions. The organizers say that Japanese researchers sometimes fade into the background at such times, mixing mainly with one another.

Yoh Iwasa, a theoretical biologist at Kyushu University who helped to organize the meeting, concedes that "Japanese researchers are not used to getting involved". But he says that at the meeting they became more vocal as the week progressed, and he is optimistic about the future. "This was only the first one," he says.

A second meeting on species extinction is set for 2006, and Okazaki meetings are also planned for September to discuss organisms living in extreme conditions, and for March next year on reproduction. Many more could follow, the organizers say, if the model proves successful. ■

Japanese researcher hits jackpot for filing patent on blue LEDs

Tokyo A Japanese materials scientist has won a record-breaking ¥20 billion (US\$190 million) award from his former employer, in recognition of his work patenting a process for making blue light-emitting diodes.

Shuji Nakamura, who is now a researcher at the University of California, Santa Barbara, won the compensation battle with the Nichia Corporation in a Tokyo district court last Friday.

The amount is the greatest sum awarded to a researcher in Japan for an invention, smashing the record set the previous day when another court ordered Hitachi to pay an ex-employee ¥163 million for helping to develop DVD technology.

The court ordered the Nichia Corporation to pay Nakamura the sum as "adequate" compensation — the vague description of the price that Japanese law says researchers must receive for their work — for the blockbuster patent that he secured in 1993.

Together the two cases represent a recognition by the Japanese courts of the frustration felt by some industrial workers at not receiving more reward for their efforts. Nakamura says that he was only paid about US\$100 for each patent he filed during his years at Nichia.

Japan industrial researchers were expected to respond enthusiastically to the award. But some companies may threaten to shift their research bases overseas to avoid similar suits in the future.



Light-hearted: Shuji Nakamura celebrates.

Butler resigns post as he awaits sentencing

Washington Thomas Butler, the infectious-disease expert convicted of fraud and breaking germ transport rules, has resigned his professorship at Texas Tech University in Lubbock. The decision may help to reduce Butler's sentence, according to one of his lawyers, Floyd Holder. "If we didn't think that, we wouldn't have done it," Holder says. In a settlement with Texas Tech, Butler will also pay the university \$300,000.

Butler was convicted on 1 December 2003 of defrauding the university in contracts he held with pharmaceutical companies. He is scheduled to be sentenced in early March for this and other offences, which include

Rotting sperm whale spills its guts

Tokyo An otherwise ordinary day in a busy Taiwanese street was interrupted last week by 60 tonnes of exploding sperm whale.

The dead whale was being delivered on the back of a truck to the laboratory of comparative anatomist Jiann-Pyng Wang at National Cheng Kung University in the western coastal city of Tainan for an autopsy. It had washed up on a nearby beach on 25 January.

The carcass exploded after gas from decomposition built up inside. Luckily for Wang, only some of the internal organs, including the intestines, fell into the street. His focus for the post mortem, the heart and lungs, were still intact.



shipping samples of plague bacterium without proper labelling.

On 20 January, the Texas medical board voted to suspend Butler's medical licence. If he is sent to prison, his medical licence will be permanently revoked. He faces potential fines and up to five years in jail.

India looks to restore sight with stem cells

Hyderabad A new clinical centre in India will use stem-cell therapy to cure blindness in people with damage to the outer surface of their eyes. President A. P. J. Abdul Kalam inaugurated the centre last week at the L. V. Prasad Eye Institute in Hyderabad.

Corneal disease and surface damage to the eye are major causes of blindness around the world. During the past two years, researchers at the institute have treated more than 180 patients by implanting cells derived from cultured adult stem cells, with 70% success. They describe the work as "the first successful large-scale application of adult stem-cell technology anywhere in the world".

The researchers say that their technique allows the reconstruction of the entire outer surface of the eye in one go — an improvement on methods that repair only the cornea.

The project was the first to be funded under the Indian government's stem-cell initiative, launched in 2001.

Pasteur Institute set to open doors in China

Paris Emerging infectious diseases such as severe acute respiratory syndrome (SARS) and bird flu will be targeted by a new Pasteur Institute centre in Shanghai. The vice-president of the Chinese Academy of Sciences, Chen Zhu, and the Pasteur's director-general, Philippe Kourilsky, signed an agreement last week in Paris to set up a branch of the prestigious French research centre, during a state visit by the Chinese President Hu Jintao.

The centre, to be funded by the Chinese government and the Pasteur Institute, will house up to 250 scientists and technicians. They will carry out teaching and research in virology, vaccines and epidemiology, with an emphasis on diseases that jump the animal-to-human species barrier. Unusually for China, the centre will have the legal status of a not-for-profit company, and will be run by independent executive and science boards. Plans for the centre are expected to be finalized in the summer.

The Shanghai centre — the first Pasteur Institute on mainland China — will add to the agency's existing network of 23 centres in 17 countries on five continents, most of them in tropical countries where the diseases under study are endemic.

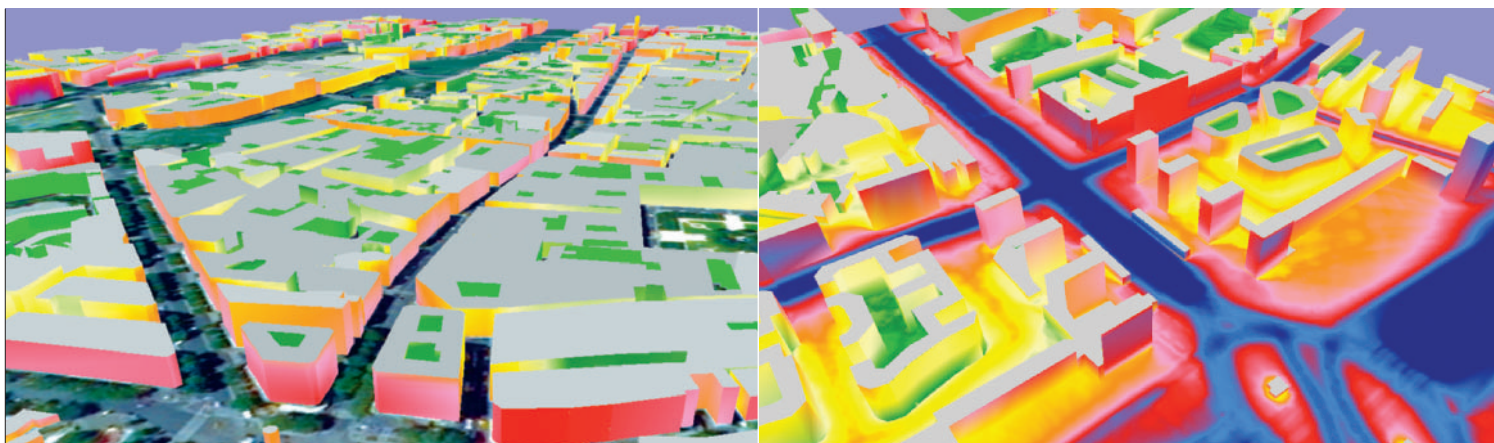
Elusive new elements make their debuts

London The periodic table has gained two new members, according to Russian and American researchers.

Fleeting glimpses were caught of elements 113 and 115 last summer at the Joint Institute for Nuclear Research in Dubna, Russia, researchers report in a paper due for publication later this month (Y. T. Oganessian *et al. Physics Reviews C*, in the press).

They fired a beam of calcium atoms at a target made of the heavy element americium. Analysis of the particles produced in the collision revealed that the new elements sprang into existence for a fraction of a second before decaying into lighter, more stable nuclei. The new elements had already been given the temporary names ununtrium and ununpentium, respectively.

Sigurd Hofmann, a nuclear physicist at the Institute for Heavy Ion Research in Darmstadt, Germany, has seen the data and says the findings look convincing. Other groups will now try to replicate the work, although Hofmann cautions that this might be difficult, as the americium target is highly radioactive and difficult to prepare.



Noise annoys: left, a part-coloured map shows noise (red) travelling up buildings; right, deafening noise of traffic entering Paris appears as deep blue.

Sound and vision

Computer simulations that paint Europe's cities in riotous colour are at the core of a bold plan to restore peace and quiet to a population driven to distraction by traffic noise. Declan Butler takes a tour.

I'm swooping down the boulevards of Paris in a virtual simulation of the city of light, where sound levels are depicted as colour. The deep blue roar of traffic washes along the streets, blazing crimson against the base of adjacent buildings, fading to orange and gold further up the façades where the noise is fainter. Orange flares of traffic echo gush down side streets, giving way to gentle shades of green in the quiet rues and parks beyond. I'm exploring one of the world's first three-dimensional (3D) computer maps of urban decibels — you might call it the city of noise.

This map hails from a bank of computers in an office of the Paris city government, and is one of the most sophisticated around. Other cities are developing their own as part of a European Union (EU) plan to tackle noise pollution. Legislation passed last year requires member states to make, by 2007, regular noise maps of all major cities, roads, railways, airports and industrial sites. These maps are a powerful new way to visualize noise pollution. Some 80 million Europeans suffer noise levels that can cause stress, sleep loss, high blood pressure and even heart attacks, according to the European Commission. Another 170 million endure noise that is just annoying.

The maps' coloured contours mean that what you see is what you hear. City planners can use them to simulate the effect of installing noise barriers or lowering the speed limit on a busy street. Changes in colour reveal instantly whether the measure has produced the desired result. The idea is to find the most cost-effective means of turning down the volume.

The most accurate way to map noise would be to place microphones every few metres throughout a city. But that would take years and cost a fortune. Instead, noise mappers resort to 'virtual microphones', each of which is a point in a computer model that reports what the sound level would be at a certain place under given circumstances. The two-dimensional Paris map has virtual microphones every 2 metres. The 3D versions (see pictures above) add microphones up the sides of buildings at 3-metre intervals vertically and every 10 metres horizontally. In total, the 3D representation of Paris contains 26 million virtual microphones.

Each microphone bases the sound level it reports on computer models of how sound from nearby noise-makers should behave. But it's impossible to model all noise sources, such as neighbours arguing and pneumatic drills. So mappers are first tackling the biggest culprit — traffic noise — which accounts for some 90% of urban noise pollution.

Virtual sound

Cars and lorries are individually much quieter than they were, but our cities are noisier because there is more traffic. Every day, about 3 million vehicles cram into the roughly 100 square kilometres of the Paris region, producing an almost inescapable din.

Each virtual microphone calculates the number of decibels emanating from engines and rolling tyres, using data on the number of vehicles passing by, their speed and the

type of road surface. The microphones also take the local topography into account. Cars climbing a hill make much more racket than when they cruise down the other side, for instance. "The flat parts of Paris were fairly straightforward, but hills like Montmartre were a lot of hassle to model," says Yann Françoise, the engineer in charge of the Paris noise map.

Input for the virtual microphones comes mainly from government records of daily traffic along the city's 1,700 km of roads.

Noise levels vary along a street, so Françoise's team divides the streets into 33,000 short stretches, and then estimates the noise on each stretch from records of the average number of cars, lorries and buses that travel them. The mappers also take into account whether the traffic is typically smooth or congested, and whether the road is paved, asphalt or cobbled.

The next step is to calculate how sound propagates. The Paris map uses software called MITHRA, developed by the French Scientific Centre for Building Physics in Paris. It models sound energy as rays and calculates how they interact with different surfaces, some materials reflecting more sound than others. MITHRA then adds up the energy of all the sound rays that hit a microphone.

Things are tricky in Paris, because its many tall buildings create what acoustical engineers call 'canyons', where noise ricochets back and forth. So although European legislation requires estimates of only the noise that reaches a virtual microphone

"The 3D simulations are breathtaking — on screen you can fly through Paris in full surround sound."



Enough to drive you mad? The din at landmarks such as the Arc de Triomphe (above) and the continual stream of traffic along the Paris ring road (right) have prompted city planners to embark on a battle against noise pollution.

directly from the source, the Paris map has to go a step further and take into account noise that bounces off a building, or the ground, or both. It takes the eight computers in Françoise's drab office at the city's Department for Protection of the Environment an entire year to produce one 3D map of Paris.

Uncharted territory

To get a good map, the input data must be of the highest quality, says Françoise. Getting decent data is the biggest problem many countries face as the 2007 deadline approaches — London, Bonn and Amsterdam are still working in only two dimensions, for example. Until new data can be gathered, many countries will be forced to cobble together huge, disparate data sets that were assembled for other purposes.

Britain has had to embark on a new national aerial-mapping survey to deal with the noise-map regulations, as its basic survey maps often do not represent buildings as individual objects, but as lines along a road. This doesn't help modellers, who need to know the heights and shapes of buildings if they are to work out how they interfere with sound.

France is fortunate in its long tradition of centralized map-making dating back to Napoleon. Consequently, most of the data needed for noise maps are stored in the National Geographical Institute in Paris. They include detailed information on the country's topography and vegetation, as well as the size and location of buildings, roads, railways and waterways. To make the noise map, Françoise's team fused these data with digitized aerial photographs of Paris, and



with a database of the precise geographical coordinates of every postal address in the city.

The results are breathtaking. On screen you fly through Paris in full surround sound. The windows of *Nature's* office in the Latin Quarter, I discover on my journey, register a quiet pale green — perfect for writing this article. Just six doors up the road, the upper floors are violet with a thumping 70 decibels.

Tests against real measurements give an error rate of ± 1 decibel, according to Françoise. "It's as good as you could hope for, so they seem to have addressed the challenges," says Kevin Keller of the environmental planning consultancy Parsons Brinckerhoff Quade & Douglas in Orange, California.

But Arne Berndt of SoundPLAN, a noise-mapping software company based in Shelton, Washington state, argues that many European models neglect some of the complex physics of noise propagation, such as interference from sound reflecting off the ground. The European Commission says that

improved models are in the works and has funded a project called Harmonoise, based in Utrecht, the Netherlands, to develop them.

Traffic calming

The Paris noise map is already proving helpful. Using data from INSEE, the national statistics agency, on how many people live at each address and on which floor, the modellers are estimating how many people would benefit from proposed noise-abatement measures, and by how much. The calculations show that about 7% of Parisians regularly endure traffic noise above 71 decibels, about as loud as a vacuum cleaner. And 46% get between 61 and 70 decibels, which can cause stress and high blood pressure on prolonged exposure.

Figures such as these have helped to support laws requiring sound insulation of façades in proportion to street noise. The city authorities are contemplating replacing asphalt with low-noise surfaces at noise hotspots. They are also negotiating with residents to close two roads in the Bois de Boulogne, the large park in western Paris, as simulations say this would create vast areas of tranquillity. More controversial for Parisians may be simulations, not yet made public, that predict that a 30-km speed limit in particular areas could restore a relatively calm 45 decibels or less — somewhere between a whisper and a conversation — to large areas of the city.

As noise maps spread through Europe, engineers and policy-makers across the Atlantic will be watching. "This is important work and we are looking forward to the results," says Robert Bernhard, co-director of the Institute of Safe, Quiet and Durable Highways at Purdue University in West Lafayette, Indiana, and secretary-general of the International Institute of Noise Control Engineering. "I wish we were doing more noise mapping in the United States."

"European legislation is far beyond anything contemplated here," adds Nicholas Miller, president of Harris Miller Miller & Hanson, an environmental-noise consultancy in Burlington, Massachusetts. The US approach is not systematic, he says — problems are only dealt with locally, usually after complaints.

How quietly Paris can be made to purr remains to be seen. As Frank Sinatra put it, metropolitans have chosen to "wake up in a city that never sleeps". A motorcycle with a broken silencer crossing Paris in the early hours can wake up a quarter of a million people. And noise maps can't remedy that. ■

Declan Butler is *Nature's* European correspondent.

Paris noise maps

♦ www.paris.fr/FR/Environnement/bruit

European noise policy

♦ <http://europa.eu.int/comm/environment/noise/home.htm#2>

Harmonoise

♦ www.harmonoise.nl

Back to the future

From reruns of a nineteenth-century experiment performed with breathtaking precision, we may gain our first glimpses of the physics that lies beyond Einstein's theories of relativity. Philip Ball reports.

A negative result is not always bad news. In 1887, the physicists Albert Michelson and Edward Morley¹ failed to detect the influence of the mysterious 'ether' — the medium through which light waves were thought to travel. As they surveyed their apparatus, held in a basement in Cleveland, Ohio, they must have been disappointed. Yet just a few years later, Albert Einstein, inspired in part by Michelson and Morley's experiment, announced his revolutionary theories of relativity² and completely reinvented our notion of space, time and gravity.

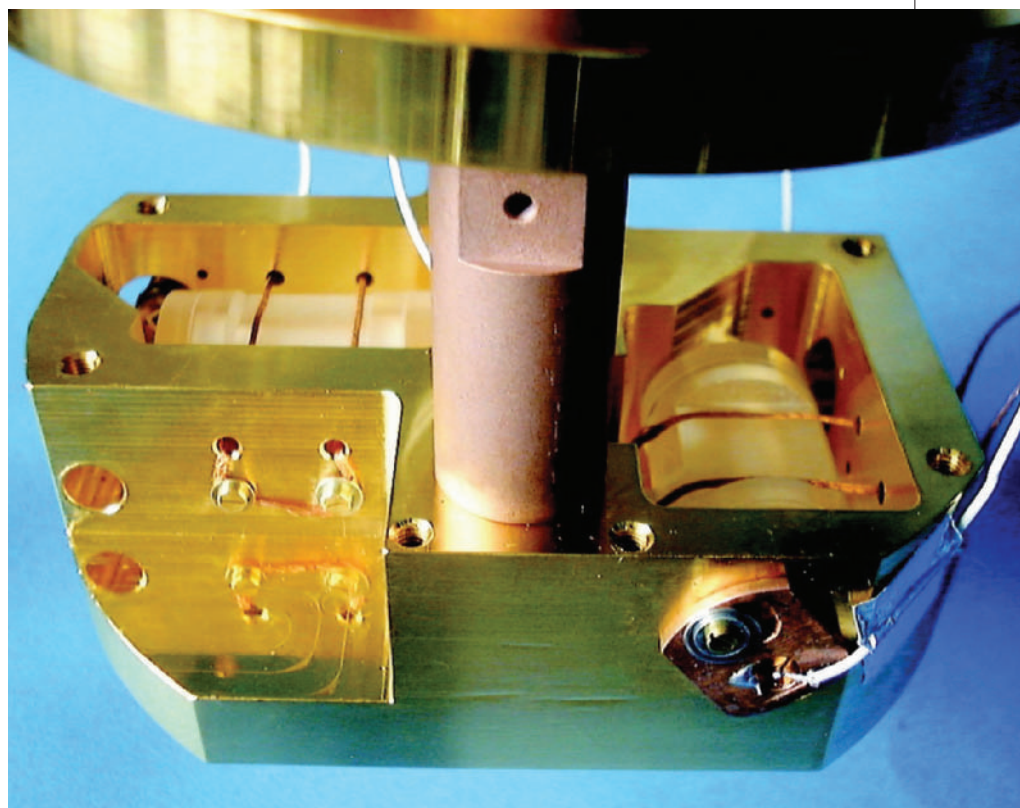
Today, experimentalists in France and Germany are rerunning Michelson and Morley's experiment with unprecedented precision^{3,4}. But this time they are not looking for the ether. By pushing this test of the constancy of the speed of light to its limits, they hope to find signs of new physics beyond Einstein's equations. If they do, they may be the first to cross the experimental threshold into an exotic new world of 'quantum gravity'. Even a negative result — essentially one that agrees with Einstein's theories — might, if it is sufficiently precise, be able to rule out or modify certain ideas about quantum gravity.

Elsewhere, the hunt for physics beyond Einstein's theories is looking out at the cosmos. A team at the University of Maryland recently put relativity to the test by studying radiation emitted by an exploding star 6,000 light years away⁵. These violent events act as natural particle accelerators, allowing scientists to probe energies far greater than any available from instruments on Earth. As the energy increases, the prospects of seeing a flaw in relativity become ever greater.

Universal theory

The possibility that experimentalists can tell which of the theorists' ideas are right or wrong is an exciting development, explains David Mattingly, a member of the Maryland team and now at the University of California, Davis. "Researchers studying quantum gravity have long lamented the 'desert' of experimental input. Now observation is catching up," he says.

This matters because we expect the laws of nature to be seamless. Yet for the past century physicists have had to get by with two incompatible theories about the way in which the Universe works. They have one



Light box: Michelson and Morley's original experimental set-up (right) and its modern-day equivalent (above) — optical resonators that measure the speed of light more accurately.

set of laws for gravity, courtesy of Einstein's general theory of relativity, and another set — quantum theory — for the other three fundamental forces of nature: electromagnetism and the strong and weak nuclear forces. This works surprisingly well most of the time, but in more extreme environments, such as inside black holes or during the early moments of the Big Bang, the two theories start coming into conflict. It would be absurd if they were not ultimately part of the same universal framework. This is why theorists are struggling to develop a quantum theory of gravity, in which the gravitational force is carried by discrete particles.

What is so special about the Michelson–Morley experiment that it could turn our notions of the Universe upside-down for a second time? Before relativity, light waves were thought to travel through the ether, a medium assumed to pervade the Universe — just as sound waves travel through air.

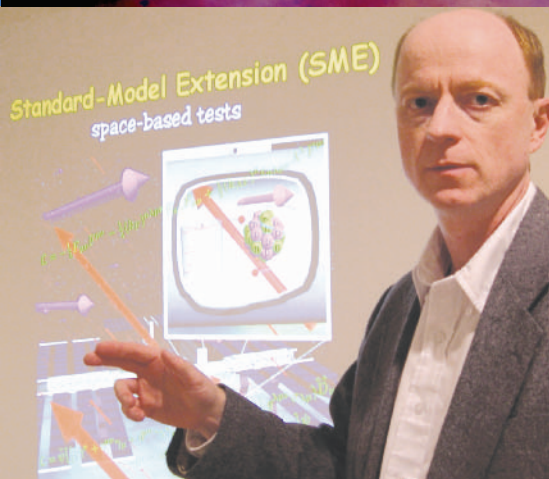


Michelson and Morley set out to detect the ether by recording the velocity of beams of light travelling in different directions. They expected to see different speeds for each beam, caused by the motion of Earth through the ether. To their surprise, they saw nothing of the sort — the speed of light remained constant in all directions.

Einstein showed why this was so. A positive result would have been in conflict with one of the cornerstones of relativity, an idea called Lorentz invariance. Named after the Dutch physicist Hendrik Lorentz, who derived it in 1904, this essentially says that the laws of physics don't change when we



The Crab nebula (above) is used to test relativity, while theorist Alan Kostelecky (left) considers light from galaxies much farther away.



switch between different reference frames, such as those moving at different speeds⁶. Lorentz also tried to explain Michelson and Morley's experiment with his idea. But instead of concluding that the ether didn't exist, he argued that the ether could define a 'universal' reference frame against which all other motions were defined.

Light standard

Einstein wasted no time in getting rid of the ether, but he kept Lorentz invariance as a fundamental feature of relativity. In essence, Einstein said there is no absolute reference frame; rather, it is the speed of light that provides an unvarying reference against which all other motions must be measured. Practically speaking, Lorentz invariance means that the result of an experiment doesn't depend on the speed or orientation of the laboratory. Einstein took care to include Lorentz invariance in relativity because he needed the laws of electromagnetism, which dictate the properties of light, to hold true everywhere and in all reference frames.

That's why the Michelson–Morley experiment remains relevant today: it is a way

of investigating to what level of accuracy the fundamental ideas of relativity hold true, without necessarily having to know anything about the new physics that lies beyond.

Some theories of quantum gravity predict that Lorentz invariance will fail over very small distances or at very high energies. So physicists are looking for signs of this failure, or 'violation', in their experiments. "The observation of Lorentz violation would be a sensitive signal for unconventional physics," says Alan Kostelecky, a theoretical physicist at Indiana University in Bloomington.

This feature of relativity is now being tested in laboratories in France and Germany using high-tech versions of the Michelson–Morley experiment. Both groups are looking for tiny changes in the time it takes a light or microwave signal to travel through crystalline structures called resonators. They use resonators made of sapphire and cooled to liquid-helium temperatures, allowing high-precision measurements.

These experiments are fundamentally similar to Michelson and Morley's in that they measure the motion of light in different reference frames. This is important because the measurement of motion changes depending on your frame of reference. For example, if a ball is thrown along a train carriage moving at 100 kilometres per hour, a passenger in the carriage sees the ball travelling at, say, 20 km h⁻¹ whereas an observer on the platform would see the ball travelling at 120 km h⁻¹.

The French team at the Paris Observatory, headed by Peter Wolf, and collaborators in Australia have measured how the motion of Earth, as it spins on its axis and orbits the Sun, affects the travel time of a light beam

circulating around a ring-shaped resonator³. This provides the researchers with two tests of Lorentz invariance that involve reference frames pointing in different directions — as in the Michelson–Morley experiment — or moving at different speeds.

So far, Einstein's equations pass both tests with flying colours — Wolf's team has detected no violation of Lorentz invariance, even though its experiment is 30 times more precise than previous tests. More recently, the team has made experimental improvements that further enhance the accuracy by a factor of two⁷. "That is about the best we can do," says Wolf. "For further improvements, you'd need a different set-up."

The German researchers, headed by Stephan Schiller at the Heinrich-Heine-University of Düsseldorf and Achim Peters at the Humboldt University of Berlin, have been using two optical resonators in their latest experiments, comparing them against each other⁴. This, they say, doubles the signal amplitude and helps to eliminate some sources of systematic error. The researchers found an upper limit for Lorentz invariance very close to that deduced by Wolf's team.

Into space

Schiller hopes that these limits will be made 1,000 times more accurate by an experiment that repeats the resonator measurements in space, on board a small satellite called OPTIS. In space, the resonators experience far less environmental vibration than on Earth, making the measurements more accurate. Schiller and Peters' team are collaborating with space researchers at the University of Bremen in Germany to develop the OPTIS project, which would also carry ultra-stable atomic clocks to provide further tests of relativity.

OPTIS has received funding from the German Space Agency, and Schiller hopes for support from the European Space Agency (ESA). "The central technology issues could be worked out in about four years," he says — but the project hangs in the balance, pending the nod from ESA. Other space-based searches for Lorentz violation are planned for the International Space Station. These include both Michelson–Morley-type experiments and ones involving atomic clocks. The earliest launch date for any of these is 2005.

What might a theory of quantum gravity look like? The electromagnetic force and the strong and weak nuclear forces are known to be transmitted by fundamental 'quantum' particles, the most familiar being the photon, for electromagnetic interactions. Physicists feel sure that the fourth fundamental force, gravity, must also have an associated quantum particle, the graviton.

But no one has yet found a way to introduce this particle in a way that fits with the picture of gravity painted by general relativity. Relativity assumes the fabric of

space-time to be smooth and continuous, whereas quantum theory wants to make it grainy. That graininess is predicted to reveal itself only at an unimaginably small scale of 10^{-35} m — called the Planck distance. This is some 20 orders of magnitude smaller than an atomic nucleus. “The Planck scale appears in almost all the different approaches to quantum gravity,” says Mattingly.

The trouble is that grainy space-time doesn’t follow the rules of relativity. In particular, many popular theories of quantum gravity suggest that, at the Planck scale, Lorentz invariance will break down. Some theories demand Lorentz violation; others seem merely to permit it. For example, the two leading candidates — string theory and loop quantum gravity — both permit Lorentz violation to varying degrees. String theory attempts to treat gravity and the other fundamental forces within the same mathematical framework, but it lacks concrete experimental predictions. Loop quantum gravity rewrites the principles of relativity in a form that is closer to quantum theory, to generate grainy space-time at the Planck scale.

High energy

These separate theories have been developed by different communities of devoted researchers over the past two decades. They have evolved into many varieties and flavours, but in the absence of any experimental data, no one can say which approach — if any — will succeed. A variation of string theory in which there are multiple membrane-like universes, or ‘brane worlds’, suggests that weak gravitons, leaking into our world from another dimension, would violate Lorentz invariance⁸. Other exotic ideas about the behaviour and structure of space-time at the Planck scale — known as noncommutative geometry and spin-foam models — also permit Lorentz violation.

Another way to catch a glimpse of the physics beyond relativity is to use the Universe itself to probe distance and energy scales we cannot hope to create on Earth. The Planck distance can be assigned an equivalent Planck energy, which physicists calculate to be about 10^{28} electron volts (eV). This is 16 orders of magnitude greater than the energies that will be available to the next generation of Earth-bound particle accelerators — such as the one



Glowing report: Peter Wolf (second from right) and his team are tracking the effects of Earth’s motion on the speed of light.

under construction at CERN, the European laboratory for particle physics, near Geneva.

Luckily, astrophysical sources can do better. Supernovae — old stars that have collapsed on themselves and then exploded — often turn into ultra-dense stars that throw out extremely energetic particles. Mattingly and his collaborators Ted Jacobson and Stefano Liberati have focused their sights on the Crab nebula, the expanding gaseous remnant of a supernova, where electrons reach energies of about 1.5×10^{15} eV.

At these energies the electrons are moving at almost the speed of light, and they emit high-energy X-rays as they get deflected by magnetic fields in the nebula — exactly as if they were in a particle accelerator. By analysing these X-rays, Mattingly and colleagues can see how fast the electrons are whizzing around. Lorentz violation would restrict the electron’s top speed, preventing it from reaching the speed of light. But the Maryland team hasn’t yet found any evidence for Lorentz violation, even though their observations allow them to extrapolate to energies seven orders of magnitude higher than the Planck energy.

Studies like this exploit not only the very high energies of astrophysical sources but also the huge distances over which their light

must travel to reach us. This allows incredibly tiny effects of Lorentz violation to be amplified into detectable ones. The idea was used by Kostelecky and his graduate student Matthew Mewes in 2001 to place the tightest bounds anyone has so far put on Lorentz violation, by looking for rotation of polarized light from very distant galaxies⁹. If Lorentz invariance is violated, the light would rotate as it travels through space — although this would be detectable only for light that has travelled over very long distances.

Although neither lab-based experiments nor astrophysical observations have directly detected Lorentz violation yet, their ‘negative’ results can already be used to probe some of the candidate theories for quantum gravity. “We are unable to rule out a whole theory,” says Mattingly, “but we can still rule out certain variants.” Calculations from loop quantum gravity, for instance, already hint at contradictions with the observations. But that doesn’t necessarily mean that every version of loop quantum gravity is wrong; the theory is still not well enough under-

stood to place that much faith in the calculations. The same is true for other ideas, such as spin-foam and brane-world models: they seem to predict too much Lorentz violation, but we can’t be sure.

Mattingly says that theorists are starting to revise their models in response to the new observations. For example, a recent paper¹⁰ already claims to have restored Lorentz invariance to quantum loop gravity. “Until recently, even this type of observational feedback to a theory of quantum gravity would have been impossible,” Mattingly says. “Theorists never used to have to revise their theories to be compatible with experiment, because there weren’t any experiments.” ■

Philip Ball is a consultant editor of *Nature*.

1. Michelson, A. A. & Morley, E. W. *Am. J. Sci.* **34**, 333 (1887).
2. Einstein, A. *Ann. Phys.* **17**, 891 (1905).
3. Wolf, P. *et al. Phys. Rev. Lett.* **90**, 060402 (2003).
4. Müller, H., Herrmann, S., Braxmaier, C., Schiller, S. & Peters, A. *Phys. Rev. Lett.* **91**, 020401 (2003).
5. Jacobson, T., Liberati, S. & Mattingly, D. *Nature* **424**, 1019–1021 (2003).
6. Lorentz, H. A. *Proc. Acad. Sci. Amsterdam* **6**, 809–831 (1904).
7. Wolf, P. *et al.* preprint at <www.arxiv.org/abs/gr-qc/0306047> (2003).
8. Burgess, C. P., Núñez, C., Quevedo, F., Zavala, I. & Tasinato, G. J. *High-Energy Phys.* **8**, 056 (2003).
9. Kostelecky, V. A. & Mewes, M. *Phys. Rev. Lett.* **87**, 251304 (2001).
10. Kozameh, C. N. & Parisi, M. F. preprint at <www.arxiv.org/abs/gr-qc/0310014> (2003).

Fertilizer 'solution' could turn local problem global

Protecting soil and water from pollution may mean releasing more greenhouse gas.

Sir—It was gratifying to see the range of well-argued responses in Correspondence (*Nature* 427, 99; 2004) to your News Feature "Fertilized to death" (*Nature* 425, 894–895; 2003). All the correspondents put forward valid points regarding the pros and cons of nitrogen fertilizer use. However, a key issue was overlooked, that of 'pollution swapping'.

Pollution of our ground and surface waters by nitrogen fertilizers poses a host of potential environmental problems, including toxic algal blooms and fish kills. Preventing nitrogen fertilizer from leaching into drainage waters, as may be achieved by no-till practices, would

therefore seem to be an obvious goal.

Here, though, we run into a real danger of what has become known as 'pollution swapping'. If the added nitrogen fertilizer is neither taken up by plants nor lost via leaching, then more of it is likely to end up as the powerful greenhouse gas nitrous oxide (A. Mosier *et al.* *Nutr. Cyc. Agroecosyst.* 52, 225–248; 1998), through the process of denitrification.

By limiting the pollution of water by nitrogen fertilizers, using so-called 'buffer strips' or strategies such as no-till farming, we may simply be swapping a relatively local pollution problem for the global problem of climate change (M. Hefting *et*

al. *J. Environ. Qual.* 32, 1194–1203; 2003).

Which of these is the more important problem depends on your perspective, and there may be other land-use strategies through which we can limit nitrogen leaching without bumping up emissions of nitrous oxide.

In the end, though, the answer for much of the developed world is likely to be a familiar one — use less fertilizer, but more efficiently.

David S. Reay

School of Geosciences, Ecology and Resource Management, University of Edinburgh, Darwin Building, Mayfield Road, Edinburgh EH9 3JU, UK

Framework reveals lack of basic-research funds

Sir—Your recent Editorial "Science needs the commission, *Sturm und Drang* and all" (*Nature* 426, 481; 2003) is most welcome. Among other things, it makes a much-needed and previously unrecognized point, namely that oversubscription by European scientists to funding offered by the European Commission's Framework programmes is "a sign of the need for European-level funding for basic research which does not exist elsewhere", not "a sign of [the programmes'] success", as naively argued by director general Achilleas Mitsos.

I would like to add a local observation. Swedish researchers have received more money through Framework programmes than the Swedish government has put into the programmes (see www.eufou.se). This is less a sign of Swedish scientists' excellence in application-writing than of an acute shortage of funding in national research councils and foundations. With an annual budget of €44 million (US\$55 million), the Swedish Medical Research Council, for example, would only be able to finance three of the Network of Excellence collaborations offered by the latest Framework programme.

Carlos F. Ibanez

Division of Molecular Neurobiology, Department of Neuroscience, Karolinska Institute, Retzius väg 8, A2:2, 171 77 Stockholm, Sweden

Working hard for the money

Sir—It is a pity that John A. Duley, in Correspondence ("Tough lessons for survival in hard academic times", *Nature* 427, 13; 2004), did not include Steven

Weinberg's fourth "golden lesson" (*Nature* 426, 389; 2003) in his own list — namely, to learn some of the history of science.

If you look back further than 25 years, it becomes clear that relatively low pay for research scientists has been a feature of the profession since its earliest days.

The junior scientists who sailed on the research vessel *Challenger* 130 years ago, for example, were paid £200 a year, roughly equivalent to £7,000 (US\$12,700) today (A. L. Rice, *Arch. Nat. Hist.* 16, 213–220; 1989), or less than 40% of the "poor salary" example given by Duley.

Albert Einstein did his most important science after hours while working at his 'real job' as a third-class patent clerk in the Bern patent office.

Gertrude Elion (who won the Nobel Prize in medicine in 1988) started working as an unpaid laboratory technician in the 1930s and only after proving herself could enjoy the "magnificent" sum of \$20 a week.

Clearly if any of these and many other individuals had chosen their research paths "according to hard-headed economics", as Duley advises, we would all be the poorer. To borrow Donna Summer's immortal words, a merchant banker "works hard for the money, so you better treat her right" — job satisfaction is not guaranteed even when the pay is good.

My own laboratory website lists three basic requirements for joining the group: an open mind, an interest in science for its own sake and a tolerable sense of humour. Disregarding the third idiosyncratic requirement, I would argue that students who fit the first two criteria should follow Steven Weinberg's advice to the letter.

Mike Fainzilber

Department of Biological Chemistry, Weizmann Institute of Science, 76100 Rehovot, Israel

Pet breeding has a long and colourful history

Sir—In his review of Tim Birkhead's entertaining book *The Red Canary* (*Nature* 425, 772; 2003), Jerry Coyne disputes the relevance of 'genetically engineering' red canaries (by crossing the birds with the South American red siskin and feeding them carotenoids) to the modern debate on transgenics. In my view, the current controversy surrounding the commercialization of GloFish, a fluorescent zebrafish with sea anemone genes (*Nature* 426, 372 and 596; 2003), could gain a useful perspective from the story of the red canary.

GloFish continues the historic tradition of genetic modification of pets that has resulted in such oddballs as peculiarly shaped dogs, hairless cats and coloured birds. The red canary is remarkable in that the red-factor gene was bred into the canary genome from a distinct species, and therefore does not differ in principle from GloFish. Members of the California Fish and Game commission, who voted against the sale of GloFish in the state and concluded that "aesthetic reasons are not sufficient justification for the genetic modification of animals", should probably endeavour to ban the sale of red canaries and other hybrid pets.

Sophien Kamoun

Department of Plant Pathology, The Ohio State University, Ohio Agricultural Research and Development Center, 1680 Madison Avenue, Wooster, Ohio 44691, USA

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

Racial realities or bombast?

When is it helpful to categorize people according to race?

Race: The Reality of Human Differences

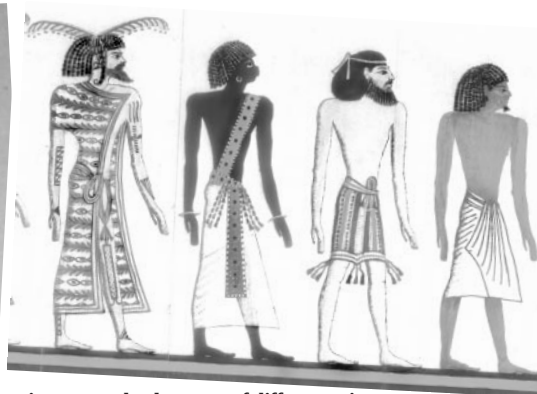
by Vincent Sarich & Frank Miele
Westview Press: 2004. 320 pp. \$27.50

Robert N. Proctor

This is a disturbing book, especially given the stature of its primary author, Vincent Sarich, as one of the founding pioneers of molecular anthropology. In 1967, in a paper with Allan Wilson, Sarich, then a graduate student at the University of California, Berkeley, used a simple protein-molecular clock to show that humans share a common ancestor with the great apes from as recently as 5 million years ago — overturning previous estimates of more 20 million years.

Here he teams up with Frank Miele, a senior editor at *Skeptic* magazine, to lament the neglect of “racial realities” by social scientists. The voice is one of loners crusading against conventional wisdom, although the book also reads as if it were a legal brief prepared for use in court to counter affirmative action. The authors’ ‘case for race’ draws heavily on contentious claims by raciologists such as Arthur R. Jensen and J. Philippe Rushton, notorious for having postulated natural racial hierarchies in intelligence, criminality, athletic performance, sexual endowment and the capacity to accumulate wealth. This is a shame, because there are good reasons to believe that certain aspects of race are very real, and that important questions of human origins, prehistoric migrations and medical therapeutics can be fruitfully addressed by properly re-examining human biovariation.

Here, though, we have an exercise in bombast and overstatement. The book begins by claiming that, throughout history, people have always had pretty much the same conception of race. Miele explains their methodology: “Vince suggested that I (Frank Miele) search the anthropology library at the University of California-Berkeley for examples of the way ancient civilizations ... described themselves and other races in their art, their literature, and their oral tradition. Did they distinguish races ... as we do today?” Based on this quick review, the authors conclude that people from ancient Egypt, Greece, Rome, India, China and the Islamic world all sorted peoples “based on the same set of characteristics — skin colour, hair form, and head shape” that we do today. As evidence, the authors display an Egyptian jug with a negroid head on one side and a rather European-looking face on the other, from which they infer a universal “common sense” reality of race.



True colours? The Greeks and Egyptians seemed to be aware of differences in race.

Flaws in this book are so numerous that it would be difficult to list them all. Long passages are quoted without attribution, and many strong claims are presented with little or no supporting evidence. We hear that children as young as three classify people on the basis of racial characteristics that they already recognize as immutable. The authors postulate “an inborn tendency to sort people into groups” and a “module” in the human brain that predisposes people to distinguish an “us” from a “them”. There are blanket assertions that “the Greeks believed in race” and that “few of us resent a rich kinsman or coethnic”. Historians will be disturbed to hear that Giordano Bruno was burned at the stake for affirming that “the earth was the center of the universe”, an assertion which, like several others in this volume, comes closer to the truth when inverted.

Stronger claims are made that border on the incendiary. In a chapter attacking affirmative action, the authors write that “a large number of white Americans harbor the suspicion that all minority members in high-status positions are there only because of affirmative action and not because of ability or achievement”. A large number? Where’s the evidence? The authors write that “All around the world downwardly mobile males who perceive themselves as being deprived of wealth, status, and especially females by up-and-coming members of a different race are ticking time bombs”. Time bombs? Again, where’s the evidence?

Sarich and Miele make similar claims in a discussion of South African plans under the apartheid regime to develop “pigmentation weapons” that would “target only black people”. After outlining how such weapons might be developed, the authors propose racial intermarriage as a “best defense” but also warn that “intermarriage, particularly of females of the majority group with males of a minority group, is the factor most likely

to cause some extremist terrorist group to feel the need to launch such an attack”. The authors add some old-style eugenics rhetoric, worrying about “plunging” birth rates in the United States and Western Europe and the “evolutionary irony lurking on the horizon” that, having conquered and colonized the world, Europeans and their descendants now risk bringing about “their own extinction” from having too few babies.

Towards the end of this book, the authors tell of a bioanthropologist colleague, Henry Harpending, who was travelling in the Kalahari and got stuck when his pick-up broke down. An ingenious bushman in the party suggested jacking up the vehicle and jump-starting it by running a rope around the back tyres and pulling. The trick worked, and the authors are clearly puzzled that an African could have come up with such an imaginative solution, given “the mean sub-Saharan IQ of 70”. Harpending’s explanation is that “bushmen are really quick and clever” and in this respect quite different from their “black African neighbors”. Sarich and Miele then ponder whether the development of agriculture may have dulled the intelligence of the continent’s former hunter-gatherers.

What I found remarkable about this story, however, is how willing the authors are to accept this low figure for the sub-Saharan IQ, based on so little supporting evidence. The authors cite an apartheid-era study of a South African high school and an ‘in press’ literature review by Rushton and Jensen, ignoring the many ways that such a sweeping and grotesque generalization could be flawed.

The authors scoff at the idea of race as a social construct, but the historical account they present is full of idealized white-and-black polarities. The authors side with Ernst Haeckel over Rudolf Virchow, Madison Grant over Franz Boas, and Carleton Coon over Ashley Montagu. There is little effort to explore which of the myriad historical

'realities' postulated for race might have alternative explanations.

I suspect that the impact of this book could be the opposite of the authors' intentions. There is much to be said for studying human genetic variability to explore questions of prehistoric ancestry and migration, and to investigate how different human populations respond to medical interventions. But the leap from these to immoderate speculations about the permanence of present-day inequalities is likely to give sceptics even more reason to question racial 'realities'.

Anthropology has a mixed history of dealings with human racial injustice (think of Carleton Coon's view that Africans became human some 200,000 years after white Europeans). The present book, so full of flim-flam and loose speculations, is more likely to re-arm than to deflate sceptics. ■

Robert N. Proctor is in the Department of History, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

Plotting the downfall of society

Historical Dynamics: Why States Rise and Fall

by Peter Turchin

Princeton University Press: 2003. 264 pp. £22.95, \$35

Joseph A. Tainter

In one of history's most extraordinary forecasts, the Greek historian Polybius in the second century BC predicted the demise of the Roman Empire some 600 years before it fell. Like others of his time, Polybius held a cyclical view of history in which societies, like biological systems, develop through growth, maturity, senescence and death. Polybius might be more celebrated today had he based his prediction on a different theory. Retrospectively, his forecast was no more challenging than anticipating the death of an ox.

Cyclical theories, like the phenomena they postulate, come and go. They have been espoused by historians from Polybius to Oswald Spengler. Serious historians have long held cyclical theories in disrepute, but now they're back, pushed in part by biologists who are accustomed to cycling or pulsing in such systems as predator-prey relationships and ecosystem development. C. S. Holling, for example, has developed a nuanced cyclical view in his 'panarchy' theory, and Kenneth Watt has explored cycling in population, resources and economics.

Beginning with the aphorism that a discipline usually matures only after it has developed mathematical theory, population biologist Peter Turchin attempts in this



Pride comes before a fall: artist Ippolito Caffi shows how the Roman Empire was left in ruins.

book to develop quantitative cyclical theory in two main areas: territorial expansion and contraction in agrarian states, and population growth and decline in relation to political stability. He has taken care to write for historians: the verbal theories and mathematics are clearly presented, and the work is thoroughly researched and erudite.

Turchin bases his 'mature' approach on the work of the fourteenth-century Arab historian Ibn Khaldun, who sought to explain why desert nomads topple North African dynasties. Ibn Khaldun argued that the founders of dynasties rule well and tax lightly. Succeeding generations, though, develop a taste for luxury, resulting in higher taxation and declining welfare. Late-phase dynasties are challenged by desert nomads who have high degrees of *asabiya*, defined as collective solidarity or a capacity for collective action. Nomads with *asabiya* topple dynasties that lack it, starting the cycle anew. Ibn Khaldun perhaps meant his theory as a critique, but Turchin takes it literally. It might be called the 'team spirit' theory of history.

In his 'metaethnic frontier theory,' Turchin proposes that areas where imperial frontiers coincide with major ethnic boundaries function as 'asabiya incubators.' High *asabiya* allows a peripheral people to expand as an old empire contracts. Turchin builds this idea into quantitative simulations of expansion and contraction in European territorial history from AD 500 to 1900. The quantification is built on ordinal scaling, judgemental assignment of values, and arbitrary cut-offs.

Like Polybius, Turchin can mimic actual outcomes despite having a dubious social theory. Metaethnic frontier theory is flawed by its primordial assumptions (for example, that ethnic groups are 'quintessential human groups' and that conflict is innate) and by failures of fact and logic. Turchin ignores problems of complexity in large societies, and asymmetrical warfare between states and

non-states. States, of course, inculcate something akin to *asabiya* in their armed forces. Ask soldiers why they fight and they will answer: "For my buddies" — rather like nomads. *Asabiya* was strangely ineffective during the centuries when North African nomads failed to expel the Carthaginians, Romans, Vandals and Byzantines. And one wonders about the ethnic solidarity of Renaissance armies that were filled with mercenaries, a matter that Turchin ignores when he simulates European territorial changes.

'Demographic-structural theory' builds on Jack Goldstone's excellent work on population growth and state breakdown, and on Turchin's own experience in population biology. This exercise quantifies how political instability and population interact. Unsurprisingly, Turchin's models show that interaction between population dynamics and a state's fiscal health produces cycles of expansion and breakdown.

This theory is on firmer ground than *asabiya*, but much of the discussion remains simplistic. In his population model, Turchin treats élites like an inert organic mass that expands and contracts with resources, ignoring the organizational aspects of hierarchy. A need for organization may raise the proportion of élite administrators regardless of resources, as in the later Roman Empire.

Turchin cites archaeological settlement data from Roman Gaul that display two peaks and troughs. The relationship between the number of archaeological sites and population is complex, as Turchin acknowledges. If this pattern reflects population oscillations, he asserts, then unchecked population growth in the first peak led to insolvency and breakdown. In fact, neither peak reflects simple population growth. The first (from the first to the second century AD) came from Romanization and settlement of veterans, the second (in the fourth century) from changes in taxation.

Quod non fecerunt historici fecerunt biologi

ART ARCHIVE/MUSEO DI ROMA/DAGLI ORTI

—biologists presume to go where historians hesitate. Social theory is a minefield, even for those experienced in it. The quantification of historical patterns is useful and important, and should have a place in historical research. But sophisticated mathematics will not improve naive social theories. ■

Joseph Tainter is at the Rocky Mountain Research Station, US Department of Agriculture Forest Service, Albuquerque, New Mexico 87102, USA.

Get connected

Me ++: The Cyborg Self and the Networked City

by William J. Mitchell

MIT Press: 2003. 312 pp. \$27.95, £18.95

Joanne Baker

We are all cyborg nomads. So argues William J. Mitchell in his new book *Me ++*, in which he explores how our intimate relationships with technology shape both our lives and the environments we build around us. *Me ++* follows on from two of Mitchell's earlier works, *City of Bits* and *E-topia*, which investigated the symbiosis of technology and design.

In *Me ++*, Mitchell describes evocatively how miniaturized and dispersed devices, and entirely dematerialized bits of digital information, fundamentally change the way we sense and relate to the world. This realization raises new questions about how we design everything from clothing to cities. Mitchell's thesis is that ever-shrinking devices will increasingly be carried along with us — becoming part of us — rather than being accessed at fixed points in space.

As we swim in a sea of digital information broadcast as electromagnetic waves, through mobile phones, computers, video screens and speakers, we augment our bodily senses through technology, effectively becoming cyborgs. The clothing, walls, buildings and cities that surround us can be thought of as skins, successive layers through which our senses operate.

According to Mitchell, the roles of these skins should be reconsidered. Microscopic sensors, machines and miniature electronic devices can easily be replicated, distributed and carried. For example, tiny temperature sensors, or digital information systems, might be woven into the threads of clothing, rather than being hung as boxes on walls. By plugging ourselves into extended webs of sensors, we might control our own personal environments and, moreover, generate new fields of interaction within entire communities, even globally.

With great clarity, Mitchell describes how traditional, localized cities developed at sites of accumulation of materials, transport and wealth. The urban structure as we know



Seeing stars: Adam Elsheimer's *Flight into Egypt* represents the Milky Way as a series of dots.

Spot the Milky Way

Ancient Mediterranean civilizations believed that the Milky Way was composed of milk spilled from the breast of a goddess. But when Galileo turned a telescope to the heavens for the first time in 1609, he showed instead that it was made up of untold numbers of individual stars. Adam Elsheimer's *Flight into Egypt* (above), which was completed in the same year, is the first depiction of the Milky Way in an oil painting.

The painting, which shows the holy family travelling in a scene illuminated by a low-hanging Moon, presents the Milky Way as a series of dots. Does this mean that some people had already understood the true nature of the galaxy, asks Francesco Bertola in his

sumptuous new book about the Milky Way, *Via Lactea* (Biblos, €39.95).

The book highlights 60 images that illustrate how artists from across the world, over more than two millennia, have represented the faint celestial arc of the Milky Way. It also includes astronomical pictures taken using ground- and space-based telescopes, particularly the Hubble.

The book's text, in English and Italian, describes the myths and legends associated with the Milky Way, as well as its science. But there is also a moral. The Milky Way, inspiration for so much art, philosophy and simple wonder, is no longer visible to most inhabitants of our light-polluted world.

Alison Abbott

it — with central business districts, coarse zoning, suburbs and ring roads — follows directly from this function.

But today, the information revolution is changing all that. In Mitchell's view, cities are no longer dominated by localized objects but by dematerialized bits. In a world where you can connect your laptop to the Internet from practically anywhere, location has become incidental, and access to information is the new currency for foraging urban nomads. "I link therefore I am," says Mitchell. Consequently, he argues, urban design should reflect this new dynamic world.

Mitchell concludes provocatively by pronouncing the death of the modernist architectural programme. Form can no longer follow function; spaces should be what we want them to be, when we want them to be. They should be flexible as our needs change. Cafes, parks, hotels and trains can all be tem-

porary workspaces for someone with a wireless laptop, so why be confined in a cubicle?

Me ++ is an exhilarating read, jam-packed with interesting facts, colourful phrases, imagery and sage insights. Some of Mitchell's ideas are not new, but he makes a powerful argument that will influence the designers of our future environment.

Mitchell raises more questions than he answers and leaves the reader wanting more. Big issues of economics, access, surveillance, privacy, security, ownership and identity are mentioned but barely explored. Similarly, there are too few glimpses of the exciting new products that may transform our lives. What form will the futuristic knowledge-based city take? Mitchell articulates the *zeitgeist*, but even he cannot predict where technology will take us — or where we will take it. ■

Joanne Baker is in the Department of Physics, Oxford University, Oxford OX1 3RH, UK.

Prisoners of the dilemma

When mathematics and biology met on a mountain.

Martin A. Nowak

Once a year, the theoretical chemist Peter Schuster took his students from the University of Vienna to a small house in the Austrian mountains. During the day we skied, of course, but in the evening the emphasis was on science. I was a first-year PhD student looking for a project. The mathematician Karl Sigmund was there and gave a talk on what was a new topic for him: the prisoner's dilemma. At the end of the talk I asked a question, and the next day Karl and I travelled back to Vienna, endlessly debating this game. In subsequent days, I visited his office and we started to do calculations. We had become prisoners of the dilemma.

I was amazed that mathematics could be used to explore the evolution of social interactions. In the prisoner's dilemma, two players have a choice between cooperation and defection. If both cooperate, they get more than if both defect. But if one cooperates while the other defects, the cooperator gets the lowest score and the defector the highest. This is the dilemma. In a single game, defection is an unbeatable strategy. The game becomes complicated and fascinating, however, if repeated. In his talk on the mountain, Karl reported that a simple strategy, tit-for-tat (TFT), was victorious in two 'world championships' conducted by Robert Axelrod. Researchers from many disciplines had submitted computer programs to play the repeated prisoner's dilemma in a round-robin tournament. The surprise winner of the first tournament was the simplest of all strategies, TFT, submitted by game theorist Anatol Rapoport. For the second tournament, others attempted to enter strategies that were superior to TFT, but Rapoport (following the maxim of British soccer leagues, 'Never change a winning team') once more fielded TFT, and again it won.

TFT cooperates on the first move and then copies whatever the opponent did in the previous round. Immediately, we noted that this strategy would be weak if errors were taken into account. Hence, we wanted to calculate a new world championship that was run by evolutionary dynamics in the presence of mistakes. Instead of a single round-robin tournament, there was a biological population of strategies evolving over many generations. Successful strategies left more offspring for the next generation. All strategies were stochastic: they had 'trembling hands' when implementing their rules.

Before encountering Karl, my speciality had been in biochemistry. I had wanted to study medicine since high school — which



An Austrian holiday led to iterated cooperation between Karl Sigmund (left) and Martin Nowak.

pleased my parents. But in the last vacation before university, I realized that my passion was to understand the natural laws of living systems, and I switched from medicine to biochemistry. My parents consulted a dictionary and discovered that biochemists work with yeast, which was also used for fermentation. They were worried about my future. At university, I found labs disappointing — experiments failed for no good reason. But theory was beautiful. You could do theory while walking through the forest or lying in the grass. Theory was not grey, but a golden tree of life.

My collaboration with Karl continued, and embarrassingly I had to tell my parents I was now working on games. The mathematical institute of the University of Vienna was in the same building as the priests' seminary and was filled with tranquillity and solemn peace. My impression was that people in those halls certainly contemplated distant and far greater worlds.

The institute had only two computers that students could use, but these were sufficient, and I was usually alone in the computer room. Karl gave all his lectures from memory. He never carried anything with him except one folded piece of paper that contained a long password and detailed instructions for how to switch on a computer. Mathematicians were in a different league, I concluded.

We often met in coffee houses, the *genius loci* of past glory. Here Kurt Goedel announced his incompleteness theorem, Ludwig Boltzmann worked on entropy, and Ludwig Wittgenstein challenged the Vienna circle. Or we walked in the Vienna Forest, visiting a meadow called 'Himmel' (Heaven), where a sign noted that here Sigmund Freud first understood the nature of dreams.

Within a year, we had conceived an evolutionary description of probabilistic strategies in the prisoner's dilemma struggling for cooperation by natural selection (see *Nature* 355, 250–253, 1992; *Nature* 364, 56–58, 1993). The prevailing paradigm, tit-for-tat, an unforgiving retaliator, was replaced by generous tit-for-tat (which always cooperates when the other person has cooperated and sometimes even when the other person has defected) and later by win–stay, lose–shift (which stays with its current choice if the score is above an aspiration level and changes otherwise). A by-product of this work was adaptive dynamics, representing a new way to look at the evolution of strategies in a continuous space. I handed in my thesis and asked Karl what to do next. He said Robert May was second to none in theoretical biology, and a few days later my train left for Oxford.

My introduction to Karl Sigmund in the Austrian mountains was the turning point that brought together mathematics and biology. But it was not the only time that my thinking or approach to science has changed course. Far from a linear path, my scientific career feels more like brownian motion, which is continuous everywhere, but differentiable nowhere. There have been many turning points that have sent me off in new directions, making me forget what I had been doing at the time.

I am no longer embarrassed to work on games. They are the generic description of evolutionary interactions among genes, cells and people. Children love games. Scientific creativity is to never stop playing. ■

Martin A. Nowak is at the Program for Evolutionary Dynamics, One Brattle Square, Harvard University, Cambridge, Massachusetts 02138, USA.

K is for koagulation

J. Evan Sadler

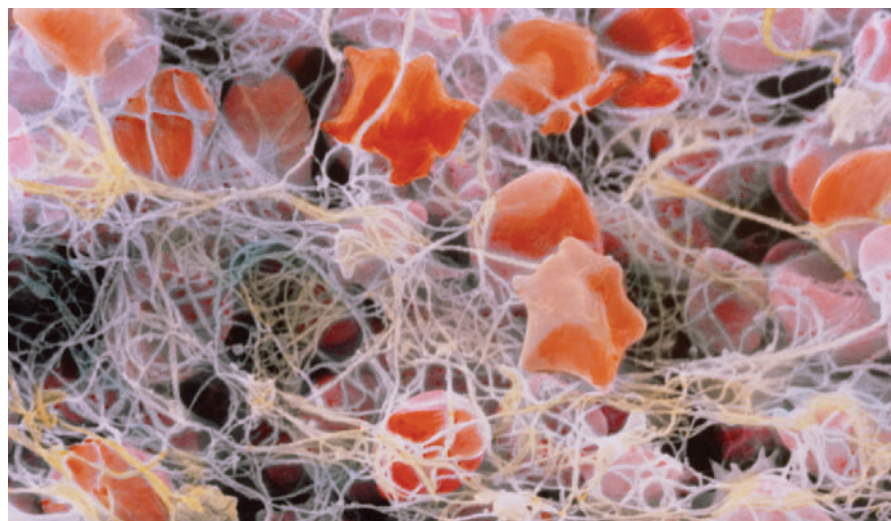
Comparative genetic linkage studies in rats, mice and humans have finally identified a key component of vitamin K metabolism that is targeted by the commonest anticoagulant drugs in use today.

Vitamins have enormous medical and economic importance. For example, the discovery of vitamin K, essential for blood clotting, led to the prevention of countless deaths from bleeding and blood clots, and facilitated the control of devastating agricultural pests. These benefits accrued despite a limited understanding of vitamin K metabolism. Vitamin K must be enzymatically activated before it can do its job, but the proteins involved in this process have resisted identification for more than 60 years. At last real progress has been made, as Rost *et al.*¹ and Li *et al.*² describe in this issue.

The 1943 Nobel Prize in Physiology or Medicine went to Denmark's Henrik Dam and Edward Doisy of the United States for their characterization of vitamin K — so named because a lack of this vitamin causes a defect in blood *koagulation* (the Scandinavian spelling). The discovery had immediate medical benefits. Before the 1940s, vitamin K deficiency often caused fatal bleeding in infants, and this has been largely eliminated by giving vitamin K to pregnant women and newborns. Meanwhile, cows in the northern United States had been bleeding to death through eating mouldy sweet clover hay. In 1940, Karl Link identified the fungal product responsible as a vitamin K antagonist. He named a potent derivative of this antagonist 'warfarin', after the Wisconsin Alumni Research Foundation, to which he assigned the patent rights.

Vitamin K deficiency impairs blood clotting by preventing the carboxylation of essential glutamic acid residues in several blood-clotting proteins. The enzyme that usually catalyses this carboxylation reaction is vitamin-K-dependent carboxylase, which uses oxygen and a reduced form of vitamin K to add a molecule of carbon dioxide to glutamic acid, producing γ -carboxyglutamic acid (Fig. 1). The modification enables clotting factors to bind calcium ions, associate with membrane surfaces and clot blood.

The other product of catalysis is vitamin K 2,3-epoxide, which is recycled to reduced vitamin K by vitamin K epoxide reductase (VKOR). VKOR activity is inhibited by warfarin. In large doses, warfarin causes bleeding and is an excellent rat poison. At just the right dose, however, it prevents the lethal blood clots that can occur in the lungs, hearts and brains of susceptible people (Fig. 1). The medical use of warfarin received a boost



P. MOTT/DEPT. ANATOMY/UNIV. "LA SAPIENZA" ROME/SPL

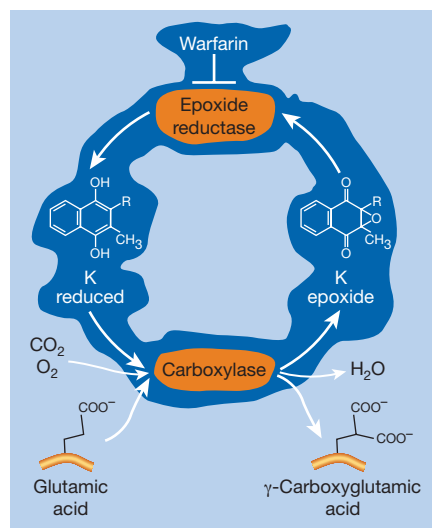


Figure 1 Left, The vitamin K cycle. Vitamin-K-dependent carboxylase uses reduced vitamin K and oxygen to add a carbon dioxide molecule to the side chain of glutamic acid in certain blood-clotting proteins, producing γ -carboxyglutamic acid and vitamin K 2,3-epoxide. Vitamin K epoxide reductase regenerates reduced vitamin K for another cycle of catalysis. Warfarin inhibits the reductase, impairing the synthesis of clotting factors and causing bleeding. A deficiency in multiple clotting factors can also be caused by mutations in either vitamin-K-dependent carboxylase or epoxide reductase. Rost *et al.*¹ and Li *et al.*² have identified the gene that encodes vitamin K epoxide reductase. Above, Close-up of a blood clot (roughly $\times 1,500$).

when US President Dwight Eisenhower was treated with it after a heart attack in 1955, and warfarin has been the most widely prescribed anticoagulant drug ever since. Some rats and mice have developed resistance to warfarin-like poisons, and some people — though very few — require extraordinarily large doses of warfarin to 'thin' their blood.

The gene that codes for vitamin-K-dependent carboxylase was cloned and characterized some time ago³. But VKOR has been purified only partially, and its gene has remained elusive. As Rost *et al.*¹ and Li *et al.*² report, a genetic approach has finally succeeded — thanks to a few key insights. First, previous work⁴ on rats with an inherited resistance to poisons had mapped the VKOR gene to rat chromosome 1, which is similar

to mouse chromosome 7. Comparisons of these rodent chromosomes with the less-similar human gene map allowed VKOR to be traced to human chromosome 10, 12 or 16. Second, some humans with a hereditary deficiency in several vitamin-K-dependent clotting factors have defective VKOR activity, and this disorder had been mapped by Fregin and colleagues⁵ to human chromosome 16. By postulating that different VKOR mutations could cause either warfarin resistance or defective clotting factors, these authors⁵ succeeded in assigning the gene to the segment of human chromosome 16 that resembles the warfarin-resistance segment on the rodent chromosomes.

Rost *et al.* and Li *et al.* therefore focused on human chromosome 16. Rost *et al.*¹ found

that there were more than 100 genes in the relevant DNA segment. To winnow the list, they turned to patients and rats with inherited defects in VKOR activity. They found that mutations in one particular gene occurred in two families with defective vitamin-K-dependent clotting factors, in four families with hereditary warfarin resistance and in numerous rat strains with resistance to warfarin-like poisons. They also identified similar genes in fish, frogs and even mosquitoes (although not in fruitflies).

Rost *et al.* show that this candidate VKOR gene encodes a small transmembrane protein that is found in the endoplasmic reticulum, a subcellular compartment. Leaving open the possibility that this protein might be only one component of a larger complex, they name it vitamin K epoxide reductase complex subunit 1 (VKORC1). The protein produced from the normal gene shows VKOR activity that is inhibited by warfarin. As expected, the mutant protein encoded by the gene from patients with a clotting-factor deficiency is inactive. Surprisingly, though, the mutant proteins from warfarin-resistant people are fairly sensitive to warfarin when tested in cultured cells; it remains to be seen how VKORC1 mutations cause warfarin resistance *in vivo*.

Li *et al.*² did not have the benefit of access to patients with VKOR mutations, but overcame this obstacle in an ingenious way. They first ruled out genes from the relevant section of chromosome 16 that code for proteins with known or predicted functions, and then selected the 13 remaining candidates that encode proteins containing potential transmembrane portions (given that VKOR seemed to be a membrane protein).

Their subsequent identification of the VKOR gene relied on the use of RNA interference. In this technique, double-stranded small interfering RNA molecules (siRNAs) are inserted into cells, where they target matching messenger RNAs for degradation. The result is that the gene encoding those messenger RNAs is effectively silenced. Li *et al.* designed siRNAs to silence each of the 13 candidate genes in human cells with a high level of VKOR activity. They found that targeting only one gene — identical to Rost and colleagues' VKORC1 gene — caused a marked reduction in VKOR activity. When expressed in insect cells, this gene produced active VKOR, which was inhibited by warfarin. As the authors point out, this siRNA strategy should be useful in other positional cloning projects for which a functional assay can be devised, even if the activity of interest depends on many gene products.

VKOR was previously thought to be a large multiprotein complex. But Li *et al.* found that expressing just one protein could confer VKOR activity on insect cells that lacked it. Although it seems a heavy burden for a small protein without obvious relation-

ship to other reductases, this molecule alone may be responsible for recycling vitamin K.

Now that we have tracked down the VKOR gene, it should be possible to identify the chemical source of the electrons that actually reduce vitamin K epoxide in living animals. Although various sulphhydryl compounds work in the laboratory, a role for other common reductase cofactors, containing nicotinamide or flavin, can now be tested. It will also be important to characterize the structure of this apparent warfarin target, perhaps leading to the identification of better vitamin K antagonists for treating blood clots or, possibly, for poisoning rats. Finally, we now know that deficiency in multiple vitamin-K-dependent clotting factors

can be caused by mutations in either of two genes — that for vitamin K carboxylase or, as shown by Rost *et al.*, for VKORC1. Yet this may not be an exhaustive list. If it is not, then linkage studies in affected families might quickly lead to the discovery of yet more components of the vitamin K cycle. ■

J. Evan Sadler is at the Howard Hughes Medical Institute, Washington University School of Medicine, St Louis, Missouri 63110, USA.
e-mail: esadler@im.wustl.edu

1. Rost, S. *et al.* *Nature* **427**, 537–541 (2004).
2. Li, T. *et al.* *Nature* **427**, 541–544 (2004).
3. Wu, S. M., Cheung, W. F., Frazier, D. & Stafford, D. W. *Science* **254**, 1634–1636 (1991).
4. Kohn, M. H. & Pelz, H. J. *Blood* **96**, 1996–1998 (2000).
5. Fregin, A. *et al.* *Blood* **100**, 3229–3232 (2002).

Planetary science

Double trouble

Joseph A. Burns

A surprising number of the icy objects in the Kuiper belt exist in pairs, or binaries. A new model proposes that these two-body systems were created through three-body interactions.

In the frigid outskirts of the Solar System, beyond the orbit of Neptune, lies the Kuiper belt. This doughnut-shaped region contains 100,000 frozen bodies, each more than 200 kilometres across. It is scarcely a decade since the first Kuiper-belt object (KBO) was sighted, but now the positions and orbits of nearly 800 are known. Sizes have been measured for a handful and, for scores more, crude colours have been detected. A few KBOs have been found to be binary — that is, two objects looping around each other while, as a pair, they circle the Sun. On page 518 of this issue, Funato *et al.*¹ present a promising explanation of how such systems might have come to exist in the outer Solar System.

When Pluto's satellite Charon was discovered in 1978, these two bodies became the first identified binary in the Kuiper belt — although this was well before the belt itself was recognized as a distinct province of our Solar System. The next pair was not found until 2001, but now a dozen binary KBOs have been pinpointed, roughly half of them by ground-based observatories and the others by the Hubble Space Telescope.

Kuiper-belt couples² seem strikingly different from other binaries³ in the Solar System. Among the known KBO pairs, the components are roughly equal in size and move on highly elongated orbits around each other, typically separated by a distance hundreds to thousands of times greater than their radii. In contrast, binary systems in the main asteroid belt, between Mars and Jupiter, are usually made up of tiny companions that are tightly bound in circular orbits around their much larger partner. The abundance and

orbital configurations of the known KBO binaries, as well as the relative sizes of the individuals themselves, provide potentially pivotal clues to how KBO binaries originated and, more generally, to how bodies accumulate and evolve in the outer Solar System.

Like the solid planets themselves, individual KBOs are believed to have formed in the early Solar System through the agglomeration of smaller bodies. So how do any KBOs end up travelling in tandem? Cataclysmic collisions probably led to the creation of Earth's Moon, the Pluto–Charon system and many asteroid companions⁴ (Fig. 1a), but the progeny of such events are born too close together and are too different in size to account for the Kuiper-belt binaries. Furthermore, each of the Kuiper-belt systems contains more angular momentum than could have been transferred in any collision that did not totally destroy and disperse the colliding bodies. The formation rate through collisions alone is, anyway, much too low to account for the total binary population of KBOs⁵.

Debris ejected in a two-body collision could, however, have been captured to yield a broad binary system had there been a third body nearby⁶. Alternatively, two passing bodies can undergo permanent mutual capture (Fig. 1b), as long as their relative energies are lowered sufficiently through the dynamical drag of small background objects or by the gravitational scattering of another large passing KBO⁷. This scheme⁴ should also generate many small, distant companions. But to produce a reasonable yield of binaries, these mechanisms^{6,7} require either many more primordial building-blocks⁶ or

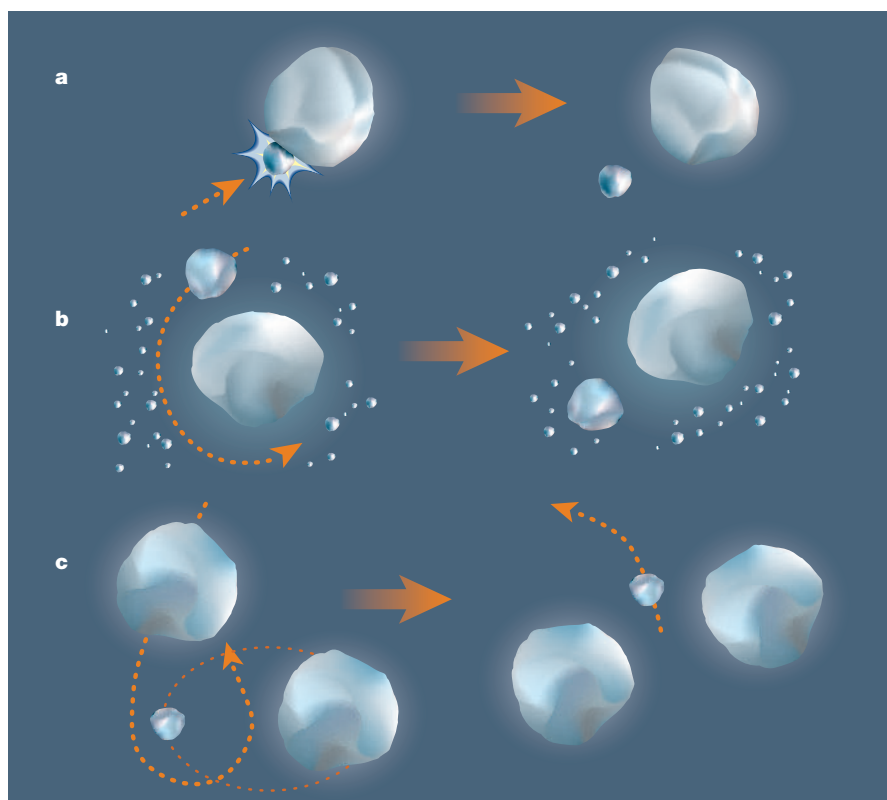


Figure 1 Doubling up. a, A binary can form through the collision of two objects. In the resulting system, one partner tends to be much smaller than the other and the bodies are relatively close to each other. In the Kuiper belt, however, binary partners are usually similar in size and farther apart. b, Alternatively, a binary may form through gravitational capture as one object passes another, with background masses tugging at the pair. In this case, partners of any relative size can link up, and again the system is unlike a Kuiper-belt binary. c, Funato *et al.*¹ propose a mechanism that combines elements of a and b. They consider an existing 'small-partner' binary, which may have formed through collision, disrupted by the subsequent gravitational capture of a larger object. The original small partner is ejected, leaving a widely separated system of two larger objects — exactly the characteristics of the known Kuiper-belt binaries.

relatively greater numbers of small bodies⁷ than expected.

Funato and colleagues' theory¹ combines aspects of both of these mechanisms^{6,7}. They suggest that objects of disparate masses initially form a close binary system through gravitational instability or through collisions. Then the pair encounters another KBO that typically has a mass similar to that of the larger of the companions (Fig. 1c). Following an elaborate gravitational dance, the two massive objects pair up and the small object is ejected from the system. This clever idea produces binaries whose attributes match those of the observed binary population of the Kuiper belt.

But caution should be the watchword when theoreticians interpret observational data before developing formation models. To begin with, binary KBOs must have been more numerous in the past⁸: there is no known mechanism that is actively creating binary systems at present, and yet binaries can be eliminated whenever energetic impacts shatter satellites and disperse the fragments. Furthermore, less catastrophic collisions and close flybys of other KBOs

might loosen the orbital embrace of pairs. These processes modify the character of any surviving pairs because they preferentially affect those binaries that contain small companions or whose orbits are weakly bound.

Moreover, because the Kuiper belt is so distant and its denizens are so small, selection effects can dramatically influence the

frequency of discoveries and mislead us about the nature of the pairs. Typical KBOs lie near the observational limit of many telescopes, and close binaries would not be resolved as two individual objects. Hence, it is no surprise that most of the binaries discovered are widely separated pairs of equally bright objects.

The binary population of the Kuiper belt had been estimated at 1–2% of the total³, but this is likely to be a severe underestimate. Two systematic searches using data from the Hubble Space Telescope, which can detect faint, close companions more efficiently than can most ground-based telescopes, have revised that estimate upwards: one² finds that $(4 \pm 2)\%$ of KBOs are binaries, the other $(7 \pm 5)\%$ (ref. 9). The latter study has been combined with a search¹⁰ using adaptive optics on the world's largest telescope, Keck II, in Hawaii. The conclusion reached is that about 10% of KBOs are binary, but that there are no wide binaries with faint companions. These are probably still lower limits, as some closely bound or very faint objects would have gone undetected.

The orbits of this increasing collection of binaries must now be determined. This information will be crucial in working out the likely origins of these objects, and in confirming the degree to which the mechanism proposed by Funato *et al.*¹ is part of the picture.

Joseph A. Burns is in the Departments of Theoretical and Applied Mechanics, and of Astronomy, Cornell University, Ithaca, New York 14853, USA.
e-mail: jab16@cornell.edu

1. Funato, Y., Makino, J., Hut, P., Kokubo, E. & Kinoshita, D. *Nature* **427**, 518–520 (2004).
2. Noll, K. *Earth Moon Planets* (in the press).
3. Merline, W. J. *et al.* in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 289–312 (Univ. Arizona Press, Tucson, 2003).
4. Michel, P., Benz, W., Tanga, P. & Richardson, D. C. *Science* **294**, 1696–1699 (2001).
5. Stern, S. A. *Astron. J.* **124**, 2300–2304 (2002).
6. Weidenschilling, S. J. *Icarus* **160**, 212–215 (2002).
7. Goldreich, P., Lithwick, Y. & Sari, R. *Nature* **420**, 643–646 (2002).
8. Petit, J.-M. & Mousis, O. *Icarus* (in the press); preprint at <http://arxiv.org/abs/astro-ph/0305156> (2003).
9. Trujillo, C. A. & Brown, M. E. *IAU Circ. No. 7787* (2002).
10. Schaller, E. L. & Brown, M. E. *Bull. Am. Astron. Soc.* **35**, 993 (2003).

Cell division

Guardian spirit blesses meiosis

Robin Allshire

During egg and sperm production, the two copies of a duplicated chromosome must be bound together until it is time for their separation. A protein that protects this chromosomal glue has now been discovered.

It is estimated that about 20% of human eggs have an abnormal number of chromosomes, and it is well known that the incidence of fetuses with three copies of some chromosomes — rather than the usual two — increases markedly with the age of the mother^{1,2}. The fact that human

eggs can arrest in the early stages of their creation for up to 45 years is clearly a factor. For this entire period, to prevent abnormal chromosome numbers, the two copies of a duplicated chromosome ('sister chromatids') need to remain tethered to each other, and the chromosome needs to be

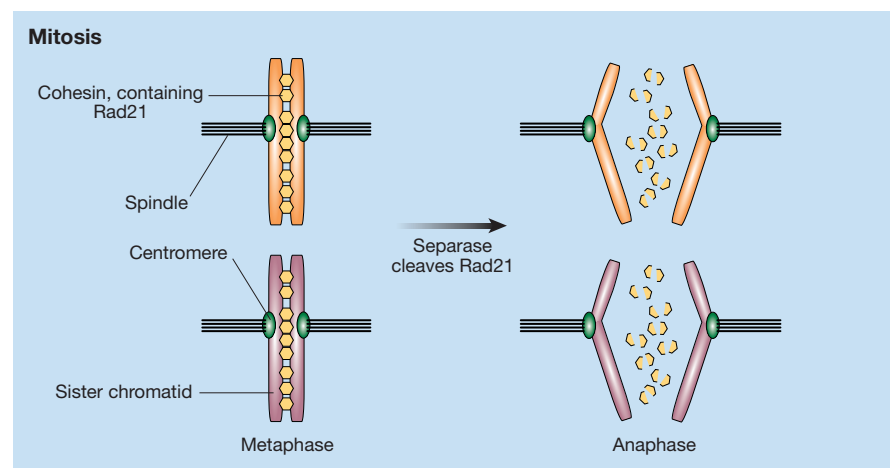


Figure 1 Mitosis. After chromosomes are duplicated, producing pairs of sister chromatids, the chromatids are kept together by a cohesin complex containing the Rad21 protein. When the centromeres attach to the spindle during metaphase, pulling forces generate tension, which activates the separase enzyme. This cleaves Rad21, resulting in the release of cohesin and allowing sister chromatids to move to opposite spindle poles during anaphase.

connected to its opposite number, until they are told to separate. On page 510 of this issue, Kitajima *et al.*³ reveal how the molecular tether between sister chromatids is kept in place.

There are two ways in which cells can multiply: mitosis and meiosis. Mitosis occurs in all the body's tissues, and generates more-or-less identical daughter cells; it serves the purpose of, for example, replacing cells lost through general wear and tear. Meiosis occurs only in reproductive tissue and generates reproductive cells.

The starting point for both mitosis and meiosis is a cell with two copies (homologues) of each chromosome, one from the mother and one from the father. These chromosomes are duplicated, producing homologous pairs of sister chromatids. During mitosis (Fig. 1), the two sister chromatids of each pair are pulled apart to opposite poles of the cell, and the cell splits into two. So, each daughter cell again has two copies of each chromosome.

Meiosis, by contrast, is divided into two stages (Fig. 2). In meiosis I, the two sister chromatids of a pair are held together, and the maternal pair is separated from the paternal pair. In meiosis II, the sister chromatids are finally separated. The result is four cells, each with half the usual number of chromosomes. The cell's genetic complement is restored when it meets a complementary reproductive cell.

During the early stages of meiosis I, the homologous pairs of sister chromatids are held together along their arms; the two sister chromatids of a pair are also glued together at specialized regions called centromeres. This bonding is achieved in part by cohesin, a complex of cohesive proteins. It is well established that cohesion along the arms must be different from cohesion at centromeres, because later in meiosis I (at the onset of

so-called anaphase I), arm cohesion is released but centromere cohesion remains intact, allowing homologous pairs to travel to opposite poles^{4,5}. Whatever is responsible for this difference must be lost by anaphase II of meiosis II, when sister chromatids, in turn, separate⁴.

Chromosome segregation at anaphases I and II involves two regulated bursts of a

specific protein-degrading activity, called separase. This cleaves one subunit (Rec8) of cohesin, first at the arms and later at the centromeres⁵. But why is Rec8 cleaved only at the arms during anaphase I? How is the Rec8 at centromeres maintained? This is the problem that Kitajima *et al.*³ address, using fission yeast, *Schizosaccharomyces pombe*, as their organism of choice.

It was known⁵⁻⁷ that Rec8 is normally expressed only during meiosis, when it takes over from its mitotic counterpart, Rad21. But when mitotic cells are manipulated, Rec8 can largely replace Rad21 (ref. 6), hinting that the expression of Rec8 is not, by itself, enough to impose the properties of meiosis I on a normal mitotic division. So it was suggested^{4,5,8} that perhaps the difference is that there is a factor that protects centromeric Rec8 during meiosis I. For various reasons, a good candidate for this factor was the fruitfly Mei-S332 protein⁸. But this molecule fell out of favour, mainly because of a lack of apparent counterparts in other organisms.

Kitajima *et al.*³ reasoned that if a protector exists in the form of a single protein, then it might usually be expressed only during meiosis, and, if forcibly expressed with Rec8 in mitotic cells, it might protect cohesion at centromeres. This would kill the cells, because sister chromatids would be unable

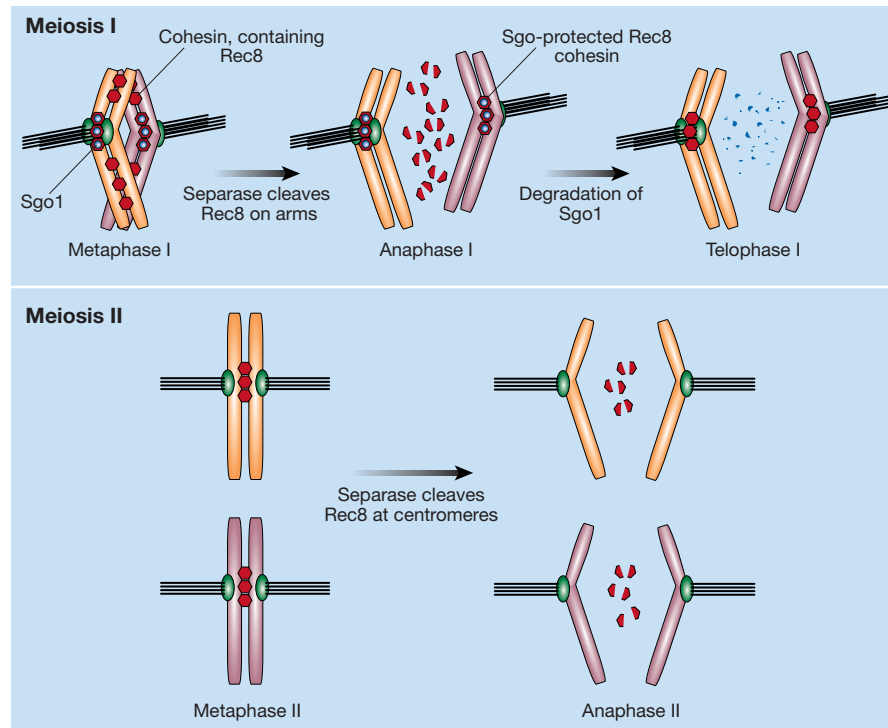


Figure 2 Meiosis. In meiosis I the centromeres of a pair of sister chromatids work in unison so that they face, and attach to, the same spindle pole. A cohesin complex containing the Rec8 protein holds sister chromatids together at centromeres and along chromosome arms. Anaphase I is triggered by cleavage of arm Rec8 by separase, allowing the homologous pairs to disengage and move to opposite poles. Kitajima *et al.*³ have found that Sgo1 protects Rec8 at centromeres during anaphase I, so that sister chromatids remain tethered. Sgo1 is degraded after anaphase I (telophase I), so that, at metaphase II, centromeric Rec8 is no longer protected from separase. Cleavage of this centromeric Rec8 at anaphase II allows sister chromatids to separate.

Nanomaterials

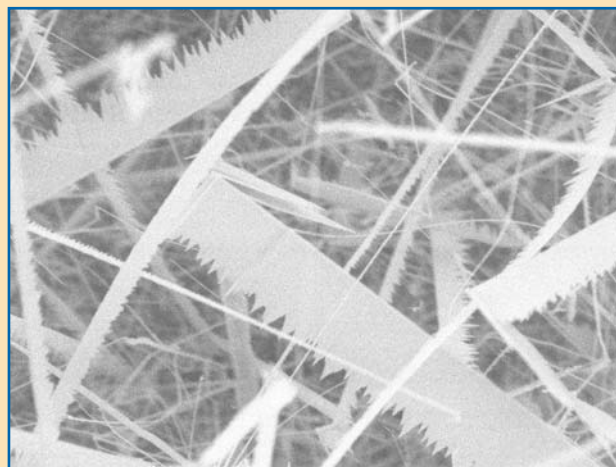
Jagged edge

It's highly debatable whether these sharp-toothed ribbons will ever be put to use as blades for nanoscale saws. But the wonder is that they form at all. Made from semiconducting cadmium selenide (CdSe), they have one edge that is atomically smooth whereas the other traces out a sawtooth shape reminiscent of a graph of random noise.

These structures, just a few micrometres wide, are produced by evaporating CdSe powder in a furnace and using a flow of nitrogen to carry the vapour to a cooler region strewn with gold nanoparticles

(C. Ma *et al.* *J. Am. Chem. Soc.* doi:10.1021/ja0395644; 2003). The gold particles act as catalysts for the regrowth of CdSe crystals as ribbon-shaped structures.

Ma *et al.* think that the asymmetry of the ribbon edges — one smooth, the other jagged — is caused by their differing chemical composition. The orientation of crystal planes in the ribbons means that their edges terminate in a layer of either cadmium or selenium. The former is more chemically reactive: in zinc oxide (which has the same crystal structure), a zinc-capped surface can act as a self-catalyst for



further growth, resulting in comb-like structures (Z. L. Wang *et al.* *Phys. Rev. Lett.* **91**, 185502; 2003). A similar process here at cadmium-

terminated edges produces lateral growth with a rough profile, which resolves itself into the sawtooth profile.

Philip Ball

to separate efficiently. So the authors screened for genes that were lethal only when expressed with Rec8 in mitotic fission yeast. They isolated one such gene, which is usually expressed only in meiosis, and named its encoded protein Sgo1, short for 'Sugoshin' — 'guardian spirit' in Japanese.

Katajima *et al.* found that expressing Sgo1 with Rec8 in mitotic cells allowed Rec8 to persist at mitotic centromeres. Moreover, Sgo1 associates with centromeres in early meiosis and remains there until anaphase I, as expected (Fig. 2). Deleting the gene results in normal chromosome segregation in meiosis I but random segregation of sister chromatids in meiosis II, because of premature loss of cohesion at anaphase I (because all Rec8 is now cleaved). The authors further show that Sgo1 and Rec8 associate in the same complex. These data are consistent with a role for Sgo1 in protecting centromeric Rec8 from separase at the onset of anaphase I.

How is Sgo1 recruited to centromeres? Here, too, Kitajima *et al.* suggest an answer. During both meiosis and mitosis, a 'checkpoint' monitors the attachment of centromeres to the spindle — the apparatus that physically separates chromosomes. This checkpoint also monitors the physical tension between sister centromeres. By doing so it ensures that sister chromatids make the correct attachments and achieve the required arrangement on the spindle⁹. Previous analyses of fission yeast¹⁰ had shown that one checkpoint component, Bub1, is required to maintain Rec8 at centromeres after anaphase I.

Kitajima *et al.* show that Bub1 is, in fact, required to recruit Sgo1 to centromeres, suggesting a connection between the protection of Rec8 and the spindle checkpoint. One

possibility is that the lack of tension between sister centromeres in meiosis I — resulting from their enforced attachment to the same spindle pole — activates Bub1, which then recruits Sgo1. Such a signal would have to differ from that normally associated with a lack of inter-centromere tension; otherwise, division would stop⁹.

Broadening their findings, Kitajima *et al.* found that fission yeast contains a protein related to Sgo1, which they call Sgo2. This protein is located at centromeres in mitotic cells, and again the association depends on Bub1. Its function is unclear, although deleting it results in viable cells with a higher rate of chromosome loss. In addition, there is a single Sgo protein, Sgo1, in budding yeast, which seems to behave similarly to fission yeast Sgo1 and Sgo2. The authors also identified proteins with regions of reasonable similarity in multicellular organisms. One of these proteins is the fruitfly Mei-S332 — so it seems that this protein's closely guarded secret is that it protects meiotic cohesion after all.

Our understanding of the role of Sgo1 in meiotic chromosome segregation has been confirmed and extended by three additional studies in both fission¹¹ and budding^{12,13} yeast. Further insight has been provided into the role of fission yeast Sgo2 that localizes with meiotic centromeres¹¹. It seems to be required in meiosis I to orient the unified sister-centromeres so that they only engage microtubules protruding from the same spindle pole. Thus, cells with Sgo2 deleted exhibit a primary defect in the segregation of chromosomes during meiosis I (ref. 11). The combined meiotic phenotype of cells lacking both Sgo1 and Sgo2 is essentially identical to that described for cells with no Bub1 (ref. 10). This is consistent with Bub1 being required to recruit these important

determiners of chromosome behaviour to meiotic centromeres.

Returning to the prevalence of chromosome-segregation defects in ageing human eggs: could one contributing factor be a premature loss of the protective influence of Sgo1 during meiosis I, resulting in aberrant chromosome segregation in meiosis II? Sgo1 itself is usually degraded during anaphase I (ref. 3), but perhaps its levels also decline stochastically with maternal age in humans. This would, however, account only for the proportion of aberrations that are caused by defective meiosis II — and these are few compared with those occurring in meiosis I (refs 1, 2). Nonetheless, the discovery of the Sgo proteins should prompt further investigations that might provide insight into defects caused by aberrant numbers of human chromosomes.

Robin Allshire is at The Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, UK.
e-mail: robin.allshire@ed.ac.uk

- Hassold, T. & Hunt, P. *Nature Rev. Genet.* **2**, 280–291 (2001).
- Champion, M. & Hawley, R. S. *Nature Cell Biol.* **4** (Suppl.), s50–s60 (2002).
- Kitajima, T. S., Kawashima, S. A. & Watanabe, Y. *Nature* **427**, 510–517 (2004).
- Rieder, C. L. & Cole, R. J. *Cell* **112**, 2607–2613 (1999).
- Petroncki, M., Siomos, M. F. & Nasmyth, K. *Cell* **112**, 423–440 (2003).
- Watanabe, Y. & Nurse, P. *Nature* **400**, 461–464 (1999).
- Klein, F. *et al.* *Cell* **98**, 91–103 (1999).
- Kerrebrock, A. W., Moore, D. P., Wu, J. S. & Orr-Weaver, T. L. *Cell* **83**, 247–256 (2004).
- Musacchio, A. & Hardwick, K. *Nature Rev. Mol. Cell Biol.* **3**, 731–741 (2002).
- Bernard, P., Maure, J. F. & Javerzat, J.-P. *Nature Cell Biol.* **3**, 522–526 (2001).
- Rabitsch, K. P. *et al.* *Curr. Biol.* (in the press).
- Katis, V., Galova, M., Rabitsch, K. P., Gregan, J. & Nasmyth, K. *Curr. Biol.* (in the press).
- Marston, A. L., Tham, W.-H., Shah, H. & Amon, A. *Science* published online doi:10.1126/science.1094220 (2004).

Ammonia transformed

Michael D. Fryzuk

Ammonia is produced industrially by combining nitrogen and hydrogen gas, catalysed over a solid iron surface. How about a catalytic reaction that could take place in solution? The first steps have now been taken.

The synthesis of ammonia from its constituent elements, nitrogen and hydrogen, ranks as one of the most important discoveries in industrial catalysis¹. It merited the award of two separate Nobel prizes — first to Fritz Haber in 1918 and then to Carl Bosch in 1931, recognizing the discovery of the process and its implementation, respectively. For more than 70 years, millions of tons of ammonia have been produced annually through the Haber–Bosch process; fertilizer made from this ammonia is estimated to be responsible for sustaining roughly 40% of the world's population and is the source for 40–60% of the nitrogen in the human body. Now a remarkable chemical transformation has been discovered (page 527 of this issue²) that is likely to have important implications for the production of ammonia.

As originally devised, the Haber–Bosch process involves the elements in their gaseous state interacting at high temperature and pressure over an activated iron surface. The iron is a heterogeneous catalyst, meaning that it is in a different phase from the reactants. The process continues to be refined, particularly now with the introduction of heterogeneous ruthenium catalysts³. A desirable improvement would be a process that operates at much lower temperatures and pressures. This could be achieved through homogeneous catalysis, involving a soluble metal complex reacting with nitrogen and hydrogen to form ammonia in a solvent. But as yet there is no known homogeneous catalyst that can effect this simple process. However, the work now reported by Pool *et al.*² suggests some elementary steps that could eventually be incorporated into a homogeneous catalytic reaction.

Pool *et al.* have traced the activation of molecular nitrogen by an organometallic complex. The complex, shown in Fig. 1a, consists of a zirconium ion to which are attached two substituted cyclopentadienyl rings (C_5Me_4H). When molecular nitrogen, or dinitrogen, is introduced, a side-on N_2 fragment is generated, bridging between two zirconium centres. This in itself is not surprising, but what happens next to the complex, in pentane solution and in the presence of hydrogen, is unprecedented: a new complex forms, in which hydrogen atoms are added to the dinitrogen bridge. Although

predicted⁴ a few years ago, this is the first observation of such a transformation. Why has it taken so long?

The answer is probably that molecular nitrogen is not that good a ligand. It is so chemically inert that even binding it to metal complexes in solution, just as Pool *et al.*² have done here, was a decades-long challenge for inorganic chemists⁵. When it can be coerced to coordinate to a metal, in many cases the interaction is so weak that other (better) ligands can easily displace the dinitrogen unit. A good example⁶ is shown in Fig. 1b: the reduction of the pentamethylated cyclopentadienyl derivative ($\eta^5-C_5Me_5$) $ZrCl_2$ with sodium amalgam (Na/Hg) in the presence of dinitrogen. The result is a complex with dinitrogen units bound end-on and bridging end-on between two zirconium centres. When exposed to hydrogen, all three dinitrogen units dissociate; hydrogen is added to each zirconium centre to generate ($\eta^5-C_5Me_5$) ZrH_2 . This is pretty typical for dinitrogen ligands⁵.

What is remarkable is how small, seemingly insignificant changes in the supporting ligands can drastically affect the outcome of such transformations. As part of an extensive study of the effects of substituents, Pool *et al.*² examined the tetramethylated version of the same cyclopentadienyl derivative — that is, one methyl group on each carbon ring is replaced by a hydrogen atom. Here, reduction of the complex with sodium amalgam in nitrogen produces a different complex, incorporating a side-on dinitrogen bridge unit, as discussed above. So, by replacing one methyl group with hydrogen, the coordination geometry of the dinitrogen unit has changed.

Now the reaction with hydrogen proceeds rapidly, without dissociation of the dinitrogen unit, to form a diazenido complex (Fig. 1a). This is worthy of comment but does have some precedent in the literature, both experimentally and theoretically. What happens next, upon thermolysis and further reaction with hydrogen, are brand new transformations: when the diazenido derivative is heated, hydrogen is lost and the central bridging N_2H_2 moiety is rearranged; further reaction in excess hydrogen results in the release of ammonia, NH_3 .

The sequence of transformations is not catalytic — this would require regeneration of the dinitrogen complex by the reaction of N_2 with ($\eta^5-C_5Me_4H$) ZrH_2 , ready to repeat

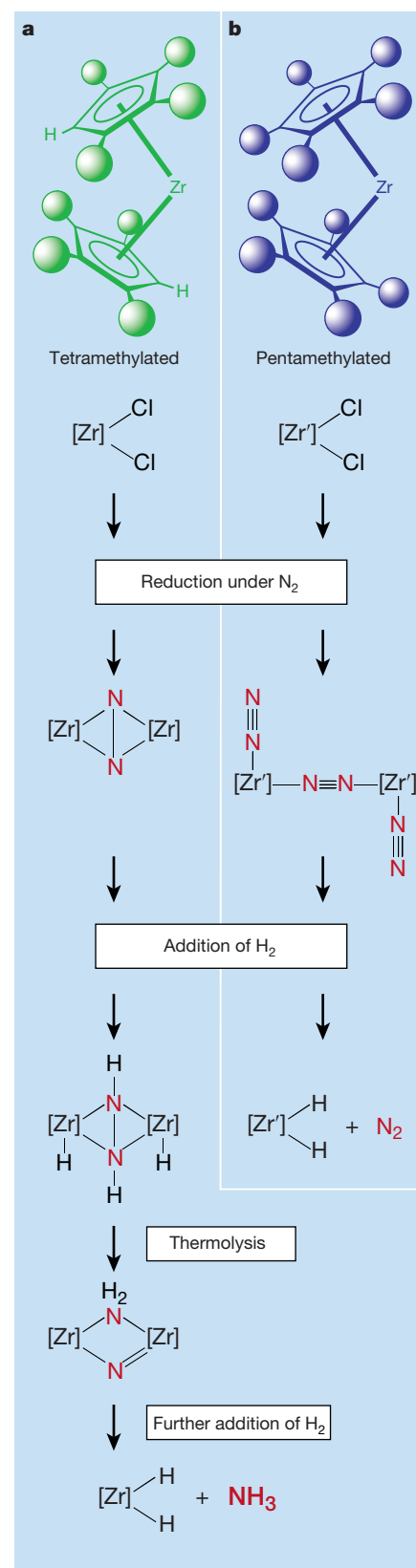


Figure 1 A new route to ammonia. a, Pool *et al.*² have traced these steps to produce ammonia from bis(tetramethylcyclopentadienyl)-zirconium dichloride (symbolized as $[Zr]Cl_2$). b, In contrast, if a slightly different ligand is used — the pentamethyl version of the complex, ($[Zr']Cl_2$) — the reaction follows a different path, and no ammonia is produced.

the process. But it is relevant, for two main reasons, to devising a homogeneous catalytic process for the production of ammonia. First, although both hydrogen and dinitrogen have been shown to bind at metal centres in solution individually and even at the same time, they have only recently been seen to interact with the formation of one N–H bond⁷. Pool and colleagues' work² suggests that these two simple molecules can react to generate a variety of NH_x ligated species. Second, and more interesting, is the remarkable effect of substituents on the ancillary ligands. The fact that changing from a pentamethylated ligand to a tetramethylated ligand can completely change how the dinitrogen unit coordinates with the metal complex, and then how it reacts with hydrogen, is truly surprising. No one could have predicted such a profound change in outcome with such a small change in ancillary ligands.

In truth, it is unlikely that any homogeneous catalytic process will ever compete on

an industrial scale with the heterogeneous Haber–Bosch reaction and its modern variants. But now that ammonia has been produced from its elements in solution, one can only begin to imagine what other kinds of transformation might be possible for molecular nitrogen. ■

Michael D. Fryzuk is in the Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, British Columbia V6T 1Z1, Canada.
e-mail: fryzuk@chem.ubc.ca

1. Tamaru, K. in *Catalytic Ammonia Synthesis* (ed. Jennings, J. R.) 1–18 (Plenum, New York, 1991).
2. Pool, J. A., Lobkovsky, E. & Chirik, P. J. *Nature* **427**, 527–530 (2004).
3. Schlögl, R. *Angew. Chem. Int. Edn* **42**, 2004–2008 (2003).
4. Basch, H., Musaev, D. G. & Morokuma, K. *Organometallics* **19**, 3393–3403 (2000).
5. Shaver, M. P. & Fryzuk, M. D. *Adv. Synth. Catal.* **345**, 1061–1076 (2003).
6. Manriquez, J. M. & Bercaw, J. E. *J. Am. Chem. Soc.* **96**, 6229 (1974).
7. Fryzuk, M. D., Love, J. B., Rettig, S. J. & Young, V. G. *Science* **275**, 1445–1447 (1997).

Ion channels

Shake, rattle or roll?

Robert O. Blaustein and Christopher Miller

Nerve transmission depends on voltage-gated ion-channel proteins, which in turn depend on the behaviour of a membrane domain called the voltage sensor. Therein lies the latest episode in a continuing story.

The propagation of electrical signals in the nervous system endows animals with moment-to-moment overall coherence. Without it, worms could not wriggle, flies could not find fruit, and we could not read News and Views. Electrical impulses coursing down the nerve axon arise from a molecular minuet in which ion-channel proteins deliver charged particles, Na^+ and K^+ , across the nerve membrane. These proteins possess a special property — voltage dependence, not known in any other macromolecule — that causes their activity to be controlled by the very voltage changes they generate. Voltage dependence arises from the transmembrane movement of charged amino acids on the protein, concomitant with opening of the ion-conducting pore. The part of the protein containing these 'gating charges', the voltage sensor, is known, but the way it moves is highly controversial. Starace and Bezannila¹ (page 548) now describe a remarkable property of the voltage sensor that sharpens our view of its transmembrane movement.

Voltage-dependent (Kv-type) K^+ channels are tetramers, with each subunit composed of six membrane-crossing α -helices, S1–S6 (Fig. 1, overleaf)². The last two of these form the K^+ -selective pore, a known structure common to all K^+ channels.

The domain that confers voltage dependence on pore opening is formed from S1–S4, with S4 bearing a regular array of positively charged amino acids, mostly arginines. The outermost three or four arginines on S4 are the gating charges that move from inside to outside to trigger pore opening. The question is, how does S4 move from the 'in' to the 'out' position?

As we see it, three classes of competing model (Fig. 2a–c, overleaf) have evolved to explain this movement, each sufficiently vague as to admit much variation in specifics. The conventional 'sliding-helix' model^{3,4} has S4 largely buried in the rest of the membrane protein, and sliding or screwing outwards like a rigid rod through a narrow, proteinaceous gasket. The 'paddle' model^{5,6}, evoked by new crystal structures, posits that S4 packs against the outer part of S3 and that this entire 'helix–turn–helix' unit, exposed to membrane lipid and positioned loosely against the periphery of the protein, flips or swivels across the membrane. Finally, the 'transporter' model^{1,7} proposes a subtle rearrangement whereby the protein moves around an essentially fixed S4. This rearrangement changes the exposure of the gating charges from the intracellular aqueous solution to the extracellular side, as in the old 'rocking-bananas' cartoons of



100 YEARS AGO

It may interest some to know that radium destroys vegetable matter. I happened to replace the usual mica plates, used to keep in the small quantity of radium in its ebonite box, with a piece of cambric, so as to permit the whole of the emanations to pass out, mica stopping the α rays. In four days the cambric was rotted away. I have replaced it now several times with the same result.

ALSO

What Mr. J. Y. Buchanan says (p. 293) about the French Academy is to me much more wonderful than the revelations of radium. It appears that there is a happy land close by where a scientific man of recognised standing can indulge in the luxury of original research, and then send in an account of his work, *not* to have it rejected by the opinion of, say, a couple of fellow-men, but actually to have it published as a right! This seems impossible. It is the encouragement of original research. Perhaps it is hopeless to expect such freedom in this stick-in-the-mud country, which is so much in love with tradition and antiquated forms.

Oliver Heaviside

From *Nature* 4 February 1904.

50 YEARS AGO

Readers of the *Bulletin of the Atomic Scientists* will have noticed that the hands of the clock which forms the distinctive coverpiece of the *Bulletin* have been moved forward so that they now stand at two minutes to midnight. This is to direct attention to the recent announcement that an atomic test involving both fission and thermonuclear reactions was conducted by the U.S.S.R. on August 12 and to warn members of all nations of the great danger of an untoward event acting as a trigger for the eruption of atomic or thermonuclear warfare. In an article in the October number of the *Bulletin* entitled "The Narrowing Way", the editor, Dr. E. Rabinowitch... considers that unless special care is taken it may only be a few more swings of the pendulum before atomic explosions strike midnight for Western civilization... Dr. Rabinowitch maintains that the Western countries are faced with the inevitable prospect of a "cold peace", precariously supported by a mutual threat of atomic and thermonuclear annihilation.

From *Nature* 6 February 1954.

solute transporters. In this last model, elegant in its minimalism, voltage dependence arises not from movement of the gating charges through the transmembrane electric field, but from the field moving around the charges. The mobile-S4 models rest on evidence variously supporting transmembrane motion on the order of 10–20 Å (refs 3, 4), whereas the transporter picture puts much weight on fluorescence experiments⁸ indicating that S4 moves less than 2 Å.

A paroxysm of recent papers^{9–12} has examined the proximity of residues near the external end of S4 to the pore domain. On channel opening, the extracellular ends of S4 and S5 come close together, and although this finding fails to distinguish among the three classes of model, it does constrain their individual depictions of the 'out' state. Pictures of the 'in' state are even murkier. The X-ray structure³ shows S4 in an intracellular position, but this structure, which everyone acknowledges is distorted by crystal-packing forces, is unrepresentative of the S1–S4 domain in a membrane. Nevertheless, electrophysiological studies^{4,6} place part of the 'in' paddle near the inside solution. In contrast, a tarantula-venom peptide binds preferentially to the 'in' state of a Kv channel, using receptor determinants on the extracellular ends of S3 and S4 (ref. 13); because the peptide inhibits from the outside, this result would rule out the paddle model if it were established that the toxin encounters its receptor directly from aqueous solution, rather than from within the membrane. But that's a big 'if'.

Against this ambiguous background, Starace and Bezanilla¹ reveal a surprising property of the 'in' configuration of S4: proton conduction. Mutation of the outermost S4 arginine to histidine produces a steady transmembrane leak of protons. This conductance behaves as though it is specifically

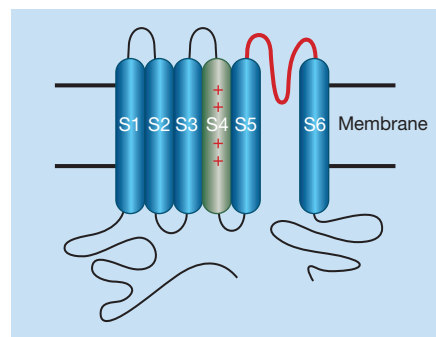


Figure 1 Composition of one subunit of a voltage-dependent K⁺ channel. A channel consists of four subunits, parts of which together form a central pore, and each subunit consists of six membrane-crossing α -helices, S1–S6. S5 and S6 are the 'pore domain' (a subunit's contribution to the pore). S1–S4 are the 'gating domain'. It is charged amino acids on S4 that are mainly responsible for voltage sensing, and movement of S4 that is the principal bone of contention.

mediated by the introduced histidine side chain, and most importantly it is present only for the voltage sensor's 'in' state. The authors estimate a unitary turnover rate of some 50,000 protons per second, and conclude that this high value implies that the histidine side chain is exposed to aqueous solution on both sides of the membrane simultaneously. From this they draw several structural inferences: that the 'in' state of the voltage sensor allows the two solutions to approach perilously close to each other, such that protons can access the side chain from both sides, and that the transmembrane voltage therefore falls across an exceedingly short distance (that of a single side chain). Declaring the paddle model inconsistent with such a picture, Starace and Bezanilla assert that voltage-sensor movement is transporter-like, as in Fig. 2c, with the outermost S4 position acting as a narrow 'gate' that separates internal and external solutions in the 'in' state.

In our view, though, it is too big a step to translate proton conductance into a unique structural image of the voltage sensor. A proton leak does not necessarily imply direct access from bulk aqueous solutions; an alternative possibility would be that the histidine residue connects to solvent via narrow crevices formed from protonatable protein groups and individual water molecules. Such pathways could act as proton conduits to the histidine side chain, as seen in proteins containing 'proton wires' as long as 15 Å (ref. 14). The distance separating the two bulk solutions could be further lengthened by about 7 Å if proton transport were also mediated by a histidine side-chain flip. Thus, the separation of internal from external solutions need not be unusually narrow. We agree that the proton current described by Starace and Bezanilla rules out a paddle completely surrounded by bilayer lipid, but a paddle in contact with the pore domain could still be consistent with these observations.

Stepping back from the fray, we should not forget that broad agreement prevails on basic issues of voltage sensing, and that the current controversy is really about fine details at the level of protein chemistry. The jury, we think, is still out, and before a firm choice of model can be made we'll need to see more experiments — and more structures.

Robert O. Blaustein is in the Molecular Cardiology Research Institute, Tufts-New England Medical Center and Department of Neuroscience, Tufts Medical School, 750 Washington Street, Boston, Massachusetts 02111, USA.

e-mail: robert.blaustein@tufts.edu
Christopher Miller is in the Department of Biochemistry, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02454, USA.
e-mail: cmiller@brandeis.edu

1. Starace, D. M. & Bezanilla, F. *Nature* **427**, 548–553 (2004).

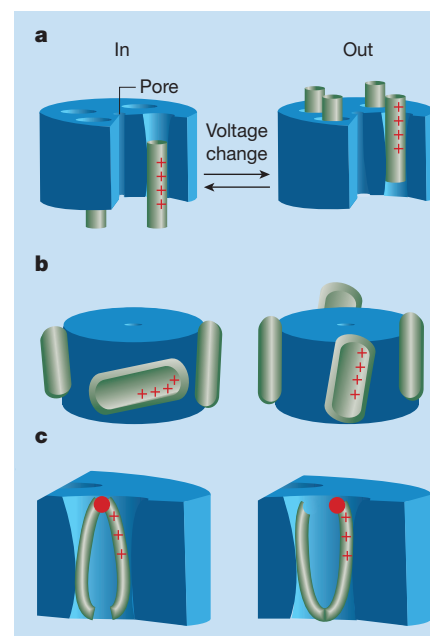


Figure 2 Models of voltage-sensor behaviour; the voltage sensors' change from the 'in' to the 'out' position triggers pore opening. a, The sliding-helix and, b, paddle models. These two diagrams show the four subunits, with the voltage-sensing components of each depicted in green. c, The transporter-like model proposed by Starace and Bezanilla¹, based on mutation of the outermost S4 amino acid from arginine to histidine (red circle), and depicted in a single subunit. According to this view of events, in its 'in' position the voltage sensor separates the internal and external solutions with a narrow gate that produces a highly focused electric field.

- Yellen, G. *Nature* **419**, 35–42 (2002).
- Yang, N., George, A. L. & Horn, R. *Neuron* **16**, 113–122 (1996).
- Larsson, H. P., Baker, O. S., Dhillon, D. S. & Isacoff, E. Y. *Neuron* **16**, 387–397 (1996).
- Jiang, Y. et al. *Nature* **423**, 33–41 (2003).
- Jiang, Y., Ruta, V., Chen, J., Lee, A. & MacKinnon, R. *Nature* **423**, 42–48 (2003).
- Yellen, G. Q. *Rev. Biophys.* **31**, 239–295 (1998).
- Cha, A., Snyder, G. E., Selvin, P. R. & Bezanilla, F. *Nature* **402**, 809–813 (1999).
- Gandhi, C. S., Clark, E., Loomis, E., Pralle, A. & Isacoff, E. Y. *Neuron* **40**, 515–525 (2003).
- Laine, M. et al. *Neuron* **39**, 467–481 (2003).
- Broomand, A., Mannikko, R., Larsson, H. P. & Elinder, F. *J. Gen. Physiol.* **122**, 741–748 (2003).
- Neale, E. J., Elliott, D. J., Hunter, M. & Sivaprasadarao, A. *J. Biol. Chem.* **278**, 29079–29085 (2003).
- Lee, H. C., Wang, J. M. & Swartz, K. J. *Neuron* **40**, 527–536 (2003).
- Luecke, H., Schobert, B., Richter, H. T., Cartailier, J. P. & Lanyi, J. K. *J. Mol. Biol.* **291**, 899–911 (1999).

Correction

In "Dreams of a hollow future" by Luis Hueso and Neil Mathur (*Nature* **427**, 301–304; 2004), the third sentence of the figure caption was added in error by an editorial hand: the property of magnetoresistance in manganites has no connection to fuel-cell technology.

Palaeoanthropology

Relative Neanderthal judgements

Proc. Natl Acad. Sci. USA
doi:10.1073/pnas.0308085100 (2004)

Neanderthals might have been a different species from modern humans — that is, separate from our own lineage. Or the taxonomic relationship could have been much closer. In the latest examination of this perennial issue, Katerina Harvati and colleagues come down in favour of the first possibility.

Neanderthals roamed Europe between about 200,000 and 30,000 years ago. A close relationship with modern humans would imply that ancient populations from different regions around the globe contributed to our evolution. On the other hand, finding Neanderthals to be a distinct species would strengthen the claim that modern humans arose relatively recently as a new species in Africa.

Harvati *et al.* have carried out a fresh assessment of the morphological differences between Neanderthals and modern humans, using variability between living primate species as a benchmark. From measurements of skulls taken from more than 1,000 specimens across 12 primate species, they show that the difference in skull morphology between Neanderthals and modern humans is similar to that between living primate species. It is also greater than the variation between populations within a species, again suggesting that Neanderthals were a distinct species.

Lucy Odling-Smee

Cryogenics

An integrated fridge

Appl. Phys. Lett. **84**, 625–627 (2004)

Some measurements are best done in the cold. For high-precision photon detection in astronomical telescopes, for example, the sensors may need cooling to a few tenths of a kelvin. Existing refrigerators capable of reaching 0.1 K are costly and cumbersome; but A. M. Clark *et al.* now describe a microelectronic device incorporated into integrated circuitry that approaches this degree of cooling cheaply and conveniently.

The device is based on a sandwich of materials: a metal, an insulator and a superconductor. With a voltage applied to the superconductor, electrons with relatively large energies ('hot' electrons) are siphoned off from the metal through a thin layer of insulating oxide, cooling the metal electrode. The oxide layer is usually derived from the superconductor, because most metal oxides don't have the right properties. But by using manganese-doped aluminium as the metal electrode, Clark *et al.* were able to invert the

conventional structure of such devices, putting the superconductor layer on top. This means that the superconductor can be thicker than usual, making it more efficient at cooling the metal. The device can thus attain lower temperatures (0.13 K) than previously possible. And it can be readily scaled up without compromising its performance.

Phillip Ball

Cell biology

Checkpoint hijack

Proc. Natl Acad. Sci. USA **101**, 947–952 (2004)

Cell lines — populations of distinct cell types that divide indefinitely under laboratory conditions — are widely used in research. They are often made by inserting a cancer-causing gene into a cell. But lines immortalized with one such gene, that for the SV40 large T antigen, are often genetically unstable — that is, their genetic make-up changes. Marina Cotsiki *et al.* have discovered why.

When a normal cell prepares to divide, it sets up a protein spindle. Pairs of duplicated chromosomes line up along the spindle's midline. Then, if the chromosomes are perfectly aligned and attached, they separate and move to opposite sides of the cell. This process is monitored by a so-called 'checkpoint' protein, Bub1. If the chromosomes are misaligned or unattached, the checkpoint kicks in to halt cell division.

But when SV40 large T antigen is present, it binds to Bub1 and overrides the checkpoint. Misaligned or unattached chromosomes are allowed to separate, and the resulting daughter cells contain an abnormal mix of chromosomes — a hallmark of cancer.

Helen R. Pilcher

Climate modelling

Freshened up

Geophys. Res. Lett. doi:10.1029/2003GL018584 (2004)

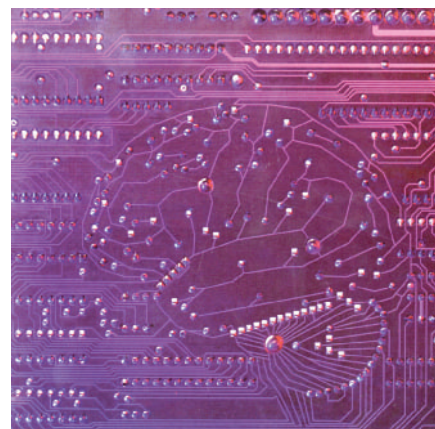
As greenhouse gases accumulate in the atmosphere, their warming effect is believed to increase the amount of fresh water released into the polar oceans from increased rainfall and melting sea ice. Many climatologists have predicted that freshening of the Arctic Ocean could disrupt the large-scale thermohaline circulation in the Atlantic — a watery 'conveyor belt' regulated by temperature and salinity — with possibly severe effects on climate. Observations from the Nordic Seas and the North Atlantic have identified such a freshening trend since the 1960s, raising questions about whether global warming is already changing ocean circulation.

Peili Wu and colleagues have now developed a climate model for the past 150 years that agrees with observations of deep-water freshening in the Nordic Seas

over the past 40 years. But their model actually predicts a short-term increase in convection in the Labrador Sea, and a more vigorous thermohaline circulation.

The authors conclude that the recently observed freshening of the deep ocean is not an early signal of climate change caused by human activities. Nevertheless, the model results are in line with the view that long-term ocean warming will result in a decline of the thermohaline circulation over the next century.

Mark Peplow



Molecular biology

Death leaves its mark

Hum. Mol. Genet. doi:10.1093/hmg/ddh065 (2004)

Many researchers analyse brain autopsy tissue in an attempt to identify genes that are working abnormally, and that might underlie psychiatric or neurological conditions. But there is an added complication: how a person dies has a profound effect on which genes were switched on in their brain.

Using a 'chip' carrying thousands of genes, J. Z. Li *et al.* compared the brains of 20 patients with depression or bipolar disorder against 20 normal controls. They found two distinct groups of active genes. Patients whose dying had been prolonged over hours or days, as a result of multi-organ failure or after coma, showed one type of gene activation profile; those who had died suddenly — from a heart attack, accident or suicide — showed another.

During prolonged illness, the brain might become starved of oxygen and essential nutrients, triggering cells to switch on 'stress response' genes for survival, the researchers suggest. Also, the brain tissue of these patients was more acidic than that of patients who had died suddenly, perhaps because their cells start to function without oxygen, churning out acidic by-products; this might also influence gene expression.

Evidently, then, researchers conducting gene-chip analyses on post-mortem brain tissue should take the patient's mode of death into account.

Helen Pearson

Getting to the bottom of a granular medium

A surprising resistance would be put up by sand grains hiding a buried treasure chest.

Penetration by an object through a dense granular medium (for example, by a finger pushing slowly into the sand on a beach) presents an interesting physics problem¹ that is closely related to issues of practical importance in soil science^{2,3}. Here we measure the penetration-resistance force for an object approaching the solid bottom boundary of a granular sample — analogous to the finger approaching a flat rock buried in the beach. We find that the penetration resistance near the boundary increases exponentially, which demonstrates the existence of an intrinsic length scale to the ‘jamming’ caused by a locally applied stress.

Our experimental apparatus consists of an actuator that pushes a flat circular plate of radius r vertically into a bucket of spherical glass beads filled to a depth of $z_{\max}$ (see Supplementary Information). To obtain constant packing throughout the sample, we filled the bucket through a tube centred in the bucket and slowly raised the tube from the bottom⁴. An in-line force cell measured the penetration force, F , as a function of the height, z , of the plate’s surface above the bottom of the bucket. The penetration force was independent of the velocity for $0.0625 \leq v \leq 2.0 \text{ mm s}^{-1}$, as expected at such low velocities⁵.

Typical measurements of $F(z)$ are shown in Fig. 1a. Far from the bottom, we expect $F(z)$ to be proportional to the ambient granular pressure, and indeed $F(z)$ is roughly linear just below the top surface (that is, for $z \approx z_{\max}$) and begins to saturate when the plate is well below the surface (as expected, because the walls bear some of the weight of the grains³). We label this intrinsic depth dependence of the penetration force (the force that is not affected by the presence of the bottom boundary) as $F_{\text{bulk}}(z)$. The effect of the bottom boundary of the bucket is apparent in Fig. 1a as a rapid increase in $F(z)$ as $z \rightarrow 0$.

To separate the effect of the bottom boundary on $F(z)$ from the intrinsic depth dependence of the penetration force in bulk material, $F_{\text{bulk}}(z)$, we acquired the functional form of $F_{\text{bulk}}(z)$ by measuring the penetration force in a bucket filled sufficiently to prevent the plate from ever approaching the bottom. As shown in Fig. 1a, by translating these data to account for different values of $z_{\max}$, we can obtain $F_{\text{bulk}}(z)$ for any value of $z_{\max}$. These results are not affected by the bottom boundary, so they provide a background measurement of the penetration force. By taking the difference between this background and the measured penetration force, we obtain the differential penetration force

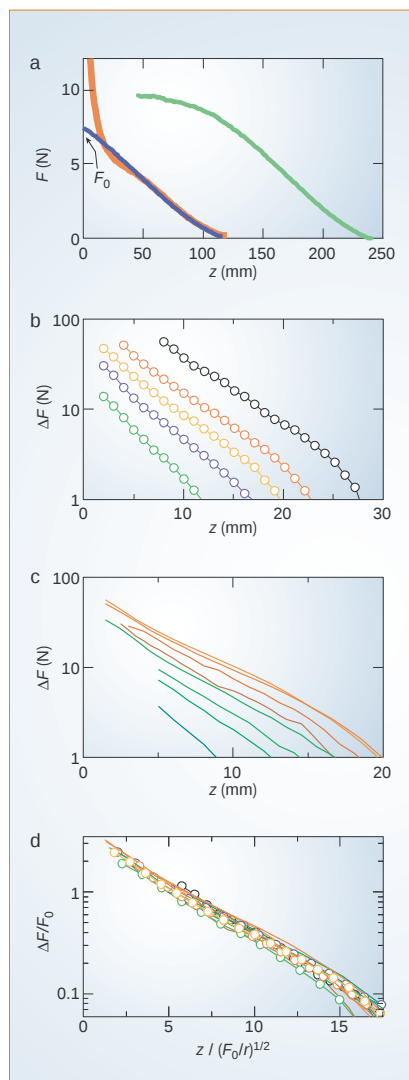


Figure 1 Penetration effects of an object approaching the solid bottom boundary of a granular sample. **a**, The penetration force, $F(z)$, for $z_{\max} = 112 \text{ mm}$ (red curve) and $z_{\max} = 230 \text{ mm}$ (green curve) was achieved by using a penetrating plate of 9.5 mm radius. $F_{\text{bulk}}(z)$ for $z_{\max} = 112 \text{ mm}$ (blue curve) was obtained by translating the $F(z)$ data acquired with $z_{\max} = 230 \text{ mm}$ so that they align with $F(z)$ data acquired with $z_{\max} = 112 \text{ mm}$. The location of F_0 , the value of $F_{\text{bulk}}(z)$ at $z = 0$, is also indicated. Note also that $F(z) < F_{\text{bulk}}(z)$ for a small regime near the bottom boundary because grains slip near the smooth bottom boundary (this effect can be altered by the bottom surface texture). All data were obtained by using monodisperse spherical glass beads (Jaygo), where $d_{\text{grain}} = 0.92 \pm 0.03 \text{ mm}$. **b**, The differential penetration force, $\Delta F(z)$, for $z_{\max} = 112 \text{ mm}$ for plate radii of 9.5, 12.7, 15.9, 19.1 and 25.4 mm (from left to right). **c**, The differential penetration force, $\Delta F(z)$, for fill heights $z_{\max}$ of 62, 82, 101, 116, 138, 153, 174 and 193 mm (from left to right along bottom) for a plate of radius $r = 12.7 \text{ mm}$. **d**, Scaled differential penetration force (see text) for data in **b** and **c**. The behaviour was exponential for rough and smooth bottom surfaces and for spherical grains as well as non-spherical sand.

$\Delta F(z) = F(z) - F_{\text{bulk}}(z)$, which quantifies the effect of the bottom surface.

Figure 1b, c shows the behaviour of $\Delta F(z)$ as the plate approaches the bottom of the bucket (that is, as $z \rightarrow 0$) for measurements varying the plate radius, r , and fill height, $z_{\max}$, respectively. The linear character of these semi-logarithmic plots indicates that $\Delta F(z)$ near the bottom is described by an exponential form, $\Delta F(z) \propto e^{-z/\lambda}$, which implies that λ is a characteristic length scale for sensing the bottom of the bucket ($\lambda \approx 5 \text{ mm}$, and is independent of grain diameter for d values of 0.3–2.3 mm). Note that this exponential behaviour must be due to collective grain motion because the bulk modulus of glass is too large to allow such a range of compression.

The results shown in Fig. 1b, c indicate that λ is affected by both the plate size and the stress state at the bottom of the bucket. Although the former can be easily measured, the latter can be quantified only through the measured quantity $F_0 = F_{\text{bulk}}$ (when $z = 0$), as indicated in Fig. 1a. The quantity F_0 is determined by the size of the granular sample, either the depth for small $z_{\max}$ or the bucket width for large $z_{\max}$ when the walls partially support the grains’ weight at the bottom. The combined data indicate that λ is proportional to $\sqrt{F_0/r}$, as shown in Fig. 1d, where a scaling of $\Delta F \rightarrow \Delta F/F_0$ and $z \rightarrow z/\sqrt{F_0/r}$ causes all of the data in Fig. 1b, c to collapse into a single characteristic exponential. This square-root dependence of λ is unexpected, although r and F_0 do correspond to the two relevant length scales for the system.

To interpret these results, it must be considered that the stress applied by a penetrating object propagates non-uniformly through force chains^{5–10} concentrated along rigid internal structures of grains known as ‘jammed’ states^{11,12}. Such states are repeatedly created and collapsed by the object’s motion. The observed exponential length scale is determined by the failure mode of the jammed state, which in turn depends on the nature of the force chains originating from the plate^{5,7,8}. The applied force from the penetrating plate in our experiments propagates through the entire granular sample to its boundary, so the observation of such a clear length scale is surprising.

Detailed modelling (as for fluids¹³) or *in situ* three-dimensional imaging would allow further characterization of the grain dynamics associated with the collapse of the jammed state. Our results indicate, however, that jamming induced by local stress can be considered as a localized phenomenon with

a well-defined extent. As jamming phenomena can be observed in diverse systems from highway traffic to foams to glasses, the possibility exists that such a length scale has more general implications that extend well beyond a finger being pushed through the sand.

Matthew B. Stone, David P. Bernstein, Rachel Barry, Matthew D. Pelc, Yee-Kin Tsui, Peter Schiffer

Department of Physics and Materials Research Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA
e-mail: schiffer@phys.psu.edu

1. Albert, R., Pfeifer, M. A., Barabási, A.-L. & Schiffer, P. *Phys. Rev. Lett.* **82**, 205–208 (1999).
2. Yu, H. S. & Mitchell, J. K. J. *Geotech. Geoenviron. Eng.* **124**, 140–149 (1998).
3. Peterson, R. W. in *Calibration Chamber Testing: PTProc. First Intl Symp. Calibration Chamber Testing* (ed. Huang, A.-B.) 315–328 (Elsevier, New York 1991).
4. Vanel, L. & Clément, E. *Eur. Phys. J. B* **11**, 525–533 (1999).
5. Geng, J. et al. *Phys. Rev. Lett.* **87**, 035506–035509 (2001).
6. Liu, C.-H. et al. *Science* **269**, 513–515 (1995).
7. Da Silva, M. & Rajchenbach, J. *Nature* **406**, 708–710 (2000).
8. Reydellet, G. & Clément, E. *Phys. Rev. Lett.* **86**, 3308–3311 (2001).
9. Bouchaud, J. P., Cates, M. E. & Claudin, P. *J. Physique I* **5**, 639–656 (1995).
10. Muegggenburg, N. W., Jaeger, H. M. & Nagel, S. R. *Phys. Rev. E* **66**, 031304–031312 (2002).
11. Liu, A. J. & Nagel, S. R. *Nature* **396**, 21–22 (1998).
12. O'Hern, C. S., Langer, S. A., Liu, A. J. & Nagel, S. R. *Phys. Rev. Lett.* **86**, 111–114 (2001).
13. Vergeles, M., Koblinski, P., Koplik, J. & Banavar, J. R. *Phys. Rev. Lett.* **75**, 232–235 (1995).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Transgenic mice

Fat-1 mice convert n-6 to n-3 fatty acids

Mammals cannot naturally produce omega-3 (n-3) fatty acids — beneficial nutrients found mainly in fish oil — from the more abundant omega-6 (n-6) fatty acids and so they must rely on a dietary supply¹. Here we show that mice engineered to carry a *fat-1* gene from the roundworm *Caenorhabditis elegans* can add a double bond into an unsaturated fatty-acid hydrocarbon chain and convert n-6 to n-3 fatty acids. This results in an abundance of n-3 and a reduction in n-6 fatty acids in the organs and tissues of these mice, in the absence of dietary n-3. As well as presenting an opportunity to investigate the roles played by n-3 fatty acids in the body, our discovery indicates that this technology might be adapted to enrich n-3 fatty acids in animal products such as meat, milk and eggs.

The *fat-1* gene of *C. elegans* encodes an n-3 fatty-acid desaturase enzyme that converts n-6 to n-3 fatty acids and which is absent in most animals, including mammals^{2,3}. We transferred this *fat-1* gene into mice and raised them alongside wild-type mice maintained on an identical diet that was high in n-6 but deficient in n-3 fatty acids. However, the fatty-acid profiles of the

two groups turned out to be quite different (Fig. 1). The tissues of wild-type animals contain polyunsaturated fatty acids that are mainly (about 98%) n-6 linoleic acid (designated an 18:2 n-6 fatty acid as it has 18 carbon atoms and 2 double bonds, one at position n-6) and arachidonic acid (AA, 20:4 n-6), with very little n-3 fatty acid (from a dietary source). By contrast, the transgenic animal tissues are rich in n-3 polyunsaturated fatty acids, including linolenic acid (18:3 n-3), eicosa-pentaenoic acid (EPA, 20:5 n-3), docosa-pentaenoic acid (DPA, 22:5 n-3) and docosahexaenoic acid (DHA, 22:6 n-3).

The concentrations of n-6 linoleic and arachidonic acids in the tissues of the transgenic mice are significantly reduced, indicating that n-6 fatty acids have been converted to n-3, causing the ratio of n-6 to n-3 to drop from 20–50 to almost 1. This n-3 enrichment at the expense of n-6 gives a balanced ratio of n-6 to n-3 and of AA/(EPA+DPA+DHA) in all of the organs and tissues tested (Table 1). Transgenic skeletal muscle contains more EPA than DHA, but DHA is the dominant n-3 fatty acid in other organs.

We have examined the tissue fatty-acid profiles in four generations of transgenic mouse lines (homozygote or heterozygote) and find consistently raised n-3 fatty acids, indicating that the transgene is functionally active *in vivo* and transmittable. The transgenic mice appear to be normal and healthy.

Efforts have been made to incorporate n-3 fatty acids into the food supply^{1,4} because of their health benefits and concern over the high n-6:n-3 ratio in Western diets. Our findings suggest a new strategy for producing food that is enriched in n-3 fatty acids from livestock carrying an n-3 desaturase trans-gene. At present, farm animals are fed fishmeal and other marine products, but this is time-consuming and costly, and is limited by the quantity of the source⁵. Production of n-3 fatty acids by the animals themselves would be a cost-effective and sustainable way of meeting the increasing demand; the ideal n-6:n-3 ratio of about 1 could be achieved by consuming foods

Table 1 Fatty-acid ratios in WT and *fat-1* mice

	n-6/n-3*		AA/(EPA+DPA+DHA)	
	WT	TM	WT	TM
Muscle	49.0	0.7	11.3	0.4
Milk†	32.7	5.7	15.7	2.5
Erythrocyte	46.6	2.9	27.0	1.6
Heart	22.8	1.8	14.3	0.9
Brain	3.9	0.8	3.6	0.7
Liver	26.0	2.5	12.5	0.9
Kidney	16.5	1.7	11.9	1.2
Lung	32.3	2.2	19.8	1.2
Spleen	23.8	2.4	17.3	1.5

Both wild-type (WT) and transgenic (TM) mice were 8-week-old females fed on the diet described in Fig. 1. AA, arachidonic acid (20:4 n-6); EPA, eicosapentaenoic acid (20:5 n-3); DPA, docosapentaenoic acid (22:5 n-3); and DHA, docosahexaenoic acid (22:6 n-3).

*The n-6:n-3 fatty-acid ratio is given by (18:2 n-6 + 20:4 n-6 + 22:4 n-6 + 22:5 n-6)/(18:3 n-3 + 20:5 n-3 + 22:5 n-3 + 22:6 n-3).

†The milk was taken from the stomach contents of 5-day-old neonatal mice born to wild-type or transgenic mothers.

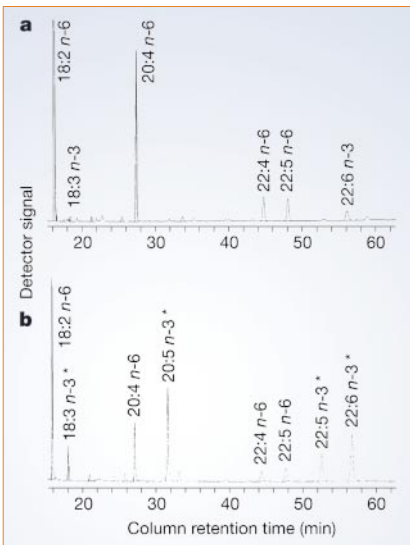


Figure 1 Partial gas chromatograph traces showing the polyunsaturated fatty-acid profiles of total lipid extracted from mouse skeletal muscle. **a**, **b**, Traces from lipid from **a**, a wild-type mouse, and **b**, a *fat-1* transgenic mouse (heterozygote). The expression vector used for microinjection contained the humanized *fat-1* sequence (with modification of codon usage) and a chicken β -actin promoter and cytomegalovirus enhancer, which allow high and broad expression of the transgene in mice^{6,7}. Both the wild-type and transgenic mice were 8-week-old females that were fed on the same diet, which was high in n-6 but low in n-3 fatty acids. The lipid profiles show that concentrations of n-6 polyunsaturated acids (18:2 n-6, 20:4 n-6, 22:4 n-6 and 22:5 n-6) are lower and levels of n-3 fatty acids (asterisks) are markedly higher in transgenic (**b**) than in wild-type (**a**) muscle. (Homozygotes and heterozygotes have a similar phenotype.)

containing this ratio and without introducing stringent dietary changes.

Our transgenic mice also offer a model for investigating the biological functions of n-3 fatty acids and the importance of the ratio of n-6:n-3 in disease prevention and treatment. Specific effects of n-3 fatty acids and of the n-6:n-3 ratio can be tested in different organs and tissues — for example, they may alter gene expression or physiological activity during the life cycle. Our mouse lines could be genetically backcrossed with mouse disease models to test the effects of n-3 fatty acids on the pathogenesis and treatment of those diseases.

Jing X. Kang*, Jingdong Wang*, Lin Wu†, Zhao B. Kang*

Departments of *Medicine and †Dermatology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA
e-mail: kang.jing@mgh.harvard.edu

1. Simopoulos, A. P. et al. (eds) *World Rev. Nutr. Diet* Vol. 83 (Basel, Karger, 1998).
2. Spychalla, J. P., Kinney, A. J. & Browse, J. *Proc. Natl Acad. Sci. USA* **94**, 1142–1147 (1997).
3. Kang, Z. B. et al. *Proc. Natl Acad. Sci. USA* **98**, 4050–4054 (2001).
4. Simopoulos, A. P. & Cleland, L. G. (eds) *World Rev. Nutr. Diet* Vol. 92 (Basel, Karger, 2003).
5. Naylor, R. L. et al. *Nature* **405**, 1017–1024 (2000).
6. Niwa, H., Yamamura, K. & Miyazaki, J. *Gene* **108**, 193–199 (1991).
7. Okabe, M. et al. *FEBS Lett.* **407**, 313–319 (1997).

Competing financial interests: declared none.

Mixing, volatile loss and compositional change during impact-driven accretion of the Earth

Alex N. Halliday

Department of Earth Sciences, ETH Zentrum, NO, Sonneggstrasse 5, Zürich, CH8092, Switzerland

The degree to which efficient mixing of new material or losses of earlier accreted material to space characterize the growth of Earth-like planets is poorly constrained and probably changed with time. These processes can be studied by parallel modelling of data from different radiogenic isotope systems. The tungsten isotope composition of the silicate Earth yields a model timescale for accretion that is faster than current estimates based on terrestrial lead and xenon isotope data and strontium, tungsten and lead data for lunar samples. A probable explanation for this is that impacting core material did not always mix efficiently with the silicate portions of the Earth before being added to the Earth's core. Furthermore, tungsten and strontium isotope compositions of lunar samples provide evidence that the Moon-forming impacting protoplanet Theia was probably more like Mars, with a volatile-rich, oxidized mantle. Impact-driven erosion was probably a significant contributor to the variations in moderately volatile element abundance and oxidation found among the terrestrial planets.

It has been proposed on the basis of new W isotopic measurements for chondrites^{1–3} that the Earth formed slightly faster than was previously thought because the original W isotopic data for carbonaceous chondrites are now known to be inaccurate by 150–200 p.p.m.. The mean life for accretion, τ , is the inverse of the time constant for exponentially decreasing growth and corresponds to the time taken for 63% of the planet to accrete^{4,5}. A τ value of ~11 Myr is obtained with continuous core formation modelling using the revised Solar System parameters¹. This new 'fast' result contrasts with the various values of 15 to 40 Myr defined by Pb isotopic data using the same models⁵. These calculations of τ assume total equilibration of incoming metal and silicate with the bulk silicate Earth (BSE) and no change in parent/daughter ratio in the total and silicate Earth during accretion. Here these assumptions are examined. It is proposed that the faster rates deduced from decay of ¹⁸²Hf to ¹⁸²W probably reflect contrasting behaviour of refractory, as opposed to volatile, elements. By modelling the data in a parallel fashion, insight into the conditions, timescales and processes of Earth accretion can be gained.

Of particular interest is whether accreting material really equilibrated isotopically with the silicate portion of the Earth—an essential tenet of accretion rate models. Some accretion and core formation simulations provide evidence against this^{6–10}. If the incoming metallic core totally fragments into very small droplets upon impact⁹ because of Rayleigh–Taylor instabilities, say, then equilibration could be relatively efficient. If, however, some of the incoming core material forms large dense masses¹⁰, then equilibration with silicate before core coagulation seems unlikely. A second issue is whether moderately volatile elements were lost during accretion¹¹. Much of the depletion in volatile elements in the terrestrial planets is thought to have occurred in the early solar nebula¹². Whether the additional dramatic differences between the Earth, Moon and Mars are the product of later losses during accretion has been less clear^{11,12}. Because Pb is moderately volatile, late changes in budgets will affect U–Pb but not Hf–W chronology. Lastly, changes in oxidation state in protoplanetary mantles can be deduced from time-integrated Hf/W ratios because of the change in core–mantle W partitioning. Therefore, W isotopes offer a finger-

print of protoplanetary environments, analogously to the recent modelling of Sr isotopes¹¹.

Realistic accretion models

Large-scale impact-driven accretion is thought to be responsible for 99% of the Earth's growth¹³. Much of the impacting material is thought to have been differentiated planets and planetesimals¹⁴, although undifferentiated volatile-rich objects probably also played a part. This complex accretion history is not well simulated by existing isotopic models. Two-stage U–Pb and Hf–W model ages have long been used to define an artificial time at which the Earth's core instantaneously formed or last equilibrated with the silicate Earth^{15–17}. Although once considered useful, such calculations, especially when based on short-lived nuclides, are of little meaning for open systems like the Earth that grew over a significant time interval^{15,13}. Dynamic simulations indicate that accretion of the Earth took about 5×10^6 years in the presence of a minimum-mass solar nebula¹⁸ or 10^7 – 10^8 years with no nebular gas¹³, rendering two-stage model ages meaningless in terms of precise rates or absolute time. This is particularly true of ¹⁸²Hf, with a half-life of 9 Myr, because late core formation would have no effect on the W isotopic composition of the Earth. The two-stage Hf–W model age for the Earth's core is 30 Myr (refs 1–3). This defines the last time, if ever, that the Earth could have globally re-equilibrated. However, there is no basis for believing that such an event occurred and, as far as W isotopes are concerned, roughly half the Earth's core could have formed much later.

A more practical and realistic approach is to assume what seems likely, that the core grew in approximately constant proportion to the Earth's mass. The isotopic data then define an accretion rate^{1,5}. There exists a strong basis for this in theory and observation. The energy from accretion and radioactive decay should result in a rapid build-up of heat, leading to widespread early melting and core formation^{8–10,19}. This is confirmed by ¹⁰⁷Pd–¹⁰⁷Ag and ¹⁸²Hf–¹⁸²W data for iron meteorites, many of which formed at pressures expected for small planetesimal cores²⁰. Most significantly, continuous core formation is consistent with the approximately constant mantle–core proportions inferred from iron depletion in silicates

from Asteroid 4 Vesta, Mars and the Earth²¹. These objects are (roughly) 1, 10 and 100% of the mass of the Earth.

Even continuous core-formation models have not addressed in detail the isotopic effects that arise with large-scale impacts such as built the Earth. Here it is assumed that accretion starts with runaway growth of objects up to 1% of the size of the Earth²². Further accretion proceeds via collisions between such differentiated objects^{13,14}. It is widely believed that the Moon formed as a result of such a collision between the proto-Earth and an impacting Mars-sized differentiated planet Theia^{6,7}, after which very little further growth occurred. Therefore, the model culminates in a '9% giant impact' (that is, an impact contributing 9% of the current mass of the Earth) which was followed by a late veneer of 0.9% of the current mass of the Earth²³. (See Fig. 1 for details.) The smooth curves in Fig. 1 show the exponentially decreasing rates of growth. The step function curves define the growth used in the isotopic calculations. The approximated τ and timing of each increment are calculated from the timing of the giant impact, t_{GI} , to achieve an overall exponentially decreasing rate of growth. The use of a step function means that the isotopic results that derive from a particular τ only approximate to that deduced from a truly continuous exponential function. Furthermore, the actual growth history is largely unknown. Neither matters for this study because it is the difference between the chronometers that is of interest.

It is assumed that the impactors and their cores grow at the same

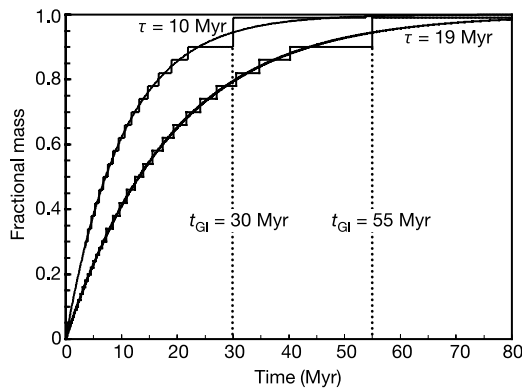


Figure 1 The change in mass fraction of the Earth as a function of time. For the model used, two accretion scenarios are shown, calculated from a giant impact at 30 and 55 Myr after the start of the Solar System. The approximated τ is the time taken to achieve 63% growth. Both this and the timing of each increment are calculated from t_{GI} to achieve an overall exponentially decreasing rate of growth for the Earth broadly consistent with dynamic simulations. The smooth curves show the corresponding exponentially decreasing rates of growth. The step function curves define the growth used in the isotopic calculations. It is assumed that the earliest stage of growth is a runaway process that builds the Earth to 1% of its current mass over $\sim 10^5$ years, as suggested by simulations²². Further growth of the Earth was dominated by stochastic collisions between these objects¹³. The overall rate of accretionary growth of the Earth may have decreased in some predictable fashion with time, but the growth events would have become more widely interspersed and larger. Therefore, the model simulates further growth by successive additions of objects of 1% of the current mass of the Earth, until it reached 10% of its current mass, then by 2% objects, up to 30% and then by 4% objects up to 90%. The Moon-forming giant impact is thought to be the last of these major collisions that affected the Earth^{6,7}. The most widely accepted scenario is that this took place when the Earth was $\sim 90\%$ of its current mass and involved an impactor planet, Theia, that was $\sim 10\%$ of the (then) mass of the Earth^{6,7}. Therefore, in the model the giant impact contributes a further 9% of the current Earth mass. There is evidence against large amounts of accretion after the giant impact. For example, significant further accretion would probably have disrupted the identical oxygen isotopic compositions of the Earth and Moon⁴⁹. In the model, a late veneer, for which some evidence exists²³, contributes an additional 0.9%. There is a final 0.1% added after this. The exact sizes of growth increments are somewhat arbitrary in this model but this is of little consequence.

absolute rate as the Earth did when of equivalent size. The isotopic evolution of metal and silicate reservoirs in the proto-Earth, and of impacting planetesimals and planets, are tracked in detail with increasing mass. The metal–silicate partitioning, the degree to which W and Pb in the cores of the impactors equilibrate with the silicate Earth, and the $^{238}\text{U}/^{204}\text{Pb}$ of the total Earth, μ_{TOT} , which may change if fractions of volatile Pb are blown-off, can all be varied at any stage. The proportion of the Moon that forms from core and silicate material from Theia, and silicate from the Earth, can also be adjusted.

Timescales for accretion from Pb isotope data for the Earth

The age of the Earth is primarily determined using Pb isotopes^{15,17,24}. The slope of the line connecting the composition of the present-day BSE with the Solar System's primordial Pb in $^{207}\text{Pb}/^{204}\text{Pb}$ versus $^{206}\text{Pb}/^{204}\text{Pb}$ space defines a model time of fractionation of U/Pb. However, in an open system like the accreting Earth, the composition relates to the accretion flux⁵. Eleven estimates^{25–35} of the Pb isotopic composition of the BSE are shown in Table 1. From these the model t_{GI} and τ necessary to achieve a growth profile like those shown in Fig. 1 have been calculated (model A, Table 1). The values of t_{GI} range between 52 and 167 Myr with τ of 18 to 57 Myr. It is hard to find a realistic mechanism for shortening these timescales to achieve agreement with Hf–W. The calculated Pb isotopic composition depends on the fractionation of $^{238}\text{U}/^{204}\text{Pb}$ for the BSE (μ_{BSE})

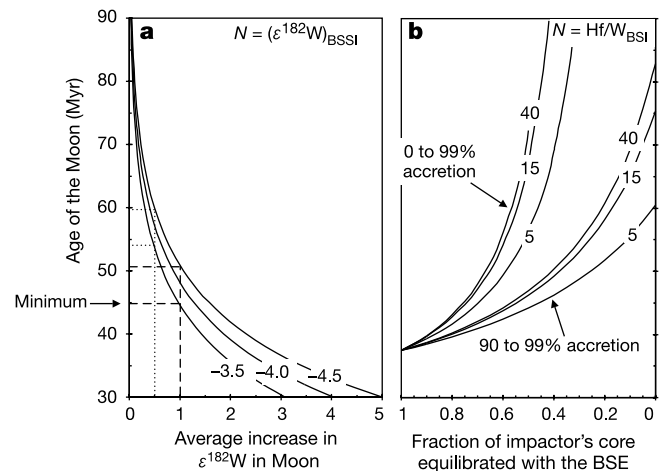


Figure 2 The age information that can be derived from W isotope data for the Earth and Moon at the present time is strongly model dependent. All calculations assume $(\text{Hf}/\text{W})_{\text{BSS}} = 1.136$, $(\text{Hf}/\text{W})_{\text{BSE}} = 15$, $(\text{Hf}/\text{W})_{\text{MOON}} = 22$, $\epsilon^{182}\text{W}_{\text{BSSI}} = -1.9$ and a half-life of hafnium-182, $\lambda^{182}\text{Hf}$, of $0.077 \times 10^{-6} \text{ yr}^{-1}$. (BSS, bulk Solar System.) In **a**, the effects on the calculated age of the Moon of not knowing the average amount of radiogenic, as opposed to inherited or cosmogenic, ^{182}W are shown. The increase in ϵW within the Moon as a function of decay of primordial ^{182}Hf , as can be deduced from currently available data, is almost certainly less than 1 unit and may be less than 0.5. The data are therefore fully consistent with independent estimates for an age of the Moon (hence giant impact) > 45 Myr after the start of the Solar System. The different curves are based on differing values³ of the $\epsilon^{182}\text{W}_{\text{BSSI}}$, which directly affects the calculated $(^{182}\text{Hf}/^{180}\text{Hf})_{\text{BSSI}}$. In **b**, the effect on calculated Hf–W timescales for the Earth's formation of incomplete mixing and equilibration of impacting core material is illustrated. This plot shows the true time of the giant impact that generates $\epsilon^{182}\text{W}_{\text{BSE}}$ of zero as a function of various levels of incomplete mixing of the impacting core material with the BSE. The lower curves are for disequilibrium during the giant impact alone. The upper curves correspond to disequilibrium during the entire accretion process up to and including the Giant Impact. The different curves for different $\text{Hf}/\text{W}_{\text{BSSI}}$ (the Hf/W in the silicate portion of the impactor) also are shown. All curves are calculated with $\epsilon^{182}\text{W}_{\text{BSSI}} = -3.5$. The effect of using more negative values³ is to yield more protracted calculated timescales for the giant impact.

relative to μ_{TOTE} during growth. The μ_{TOTE} of 0.7 used here is based on Earth's K/U ratio (ref. 15). Even if the K/U ratio of the Earth has been overestimated³⁶, a reduction only serves to increase μ_{TOTE} , extending the timescales. A 10% increase in μ_{TOTE} changes t_{GI} from >52 to >54 Myr. It has been proposed that the μ_{TOTE} is much higher and that core formation only increased μ_{BSE} by a factor of 2.5 (ref. 37). This increases the calculated timescales greatly (Table 1, model B). Incomplete mixing and lack of equilibration between the Pb of the core of Theia and the BSE during the giant impact would also increase timescales.

It is conceivable that a fraction of the Earth's Pb was lost during accretion. This effect should be small because of Earth's gravity. A comparison between the $(\text{Rb}/\text{Sr})_{\text{BSE}}$ and the time-integrated Rb/Sr of the material that made the Moon, assuming a similar starting composition, indicates a $<50\%$ loss of Rb from the Earth¹¹. Assuming $\sim 90\%$ of Earth's Pb is partitioned into the core, such loss is less likely (probably $<5\%$). In the unlikely scenario that μ_{TOTE} did change from ~ 0.35 to ~ 0.7 during the giant impact, the calculated t_{GI} would still be in the range 40 to 135 Myr (Table 1, model C).

Therefore, no likely scenarios that might have affected U–Pb preferentially eliminate the discrepancy with the Hf–W τ of 11 Myr and t_{GI} of 30 Myr. It is possible that all of the Pb estimates are incorrect, but even the three most recently published^{25–27} yield $t_{\text{GI}} > 50$ Myr and $\tau > 20$ Myr (Table 1). These timescales are similar to those deduced from I–Xe and Pu–Xe chronometry³⁸, which indicate that the last major Xe loss from the Earth's atmosphere took place 50 to 80 Myr after the start of the Solar System³⁹. It has been proposed that this was caused by atmospheric blow-off during the giant impact^{39,40}.

The age of the Moon

The age of the Moon itself provides an important test of whether such model Pb age calculations are correct. However, this is not yet well defined. Most of the precise estimates for the earliest age of the Moon are based on U–Pb and Sm–Nd and are 70 to 120 Myr (ref. 41), 50 to 70 Myr (ref. 42) and 50 to 100 Myr (ref. 43). The initial Sr isotopic composition of the Moon^{11,43} also makes sense in terms of an origin at ≥ 50 Myr because this yields a realistic time-integrated Rb/Sr for the mix of Earth and Theia of ≤ 0.07 —similar to or less than the present-day Rb/Sr of Mars, as expected¹¹.

At one time it was thought that there were well-defined W isotope variations produced by ^{182}Hf decay within the Moon, yielding model ages of ~ 55 Myr (ref. 44). A major portion of this was produced from cosmogenic ^{182}Ta (refs 45, 46). The revised average Solar System composition has been used to yield a Hf–W model age

of ~ 30 Myr for the Moon^{1,2}. This does not define the actual age of the Moon, however, because it assumes that the Moon formed from undifferentiated (with respect to Hf–W) chondritic material. There exists a great deal of evidence against this, and using the chondritic value as a reference merely provides a likely minimum time-span.

Most values of ϵW (the fractional deviation in parts per 10^4 from the $^{182}\text{W}/^{184}\text{W}$ for the BSE) values for lunar rocks that are reasonably precise ($\leq \pm 1$) after crude cosmogenic corrections are in the range zero to one^{5,45,46}. One sample has a well-defined excess of $\epsilon\text{W} \approx 1.5$ based on internal systematics⁴⁶. Some of the remaining data needed to be treated with caution until more accurate cosmogenic corrections can be applied. Nevertheless, there is a prevalence of ϵW values close to zero that has been taken as evidence that this was the initial composition still preserved in lunar reservoirs that underwent little subsequent radiogenic increase⁵. The average radiogenic increase of ϵW of the lunar mantle is probably <1.0 and may well be <0.5 . Therefore, the Hf–W age of the Moon is probably >44 and may be >54 Myr (Fig. 2a). The $\epsilon\text{W}_{\text{BSSI}}$ (where BSSI is bulk Solar System initial) is also not yet well defined and may have been lower than recently assumed³. If so, the Moon formed even later (Fig. 2a).

Tungsten isotopic disequilibrium and mixing during accretion

Given the above tentative but self-consistent evidence that the giant impact probably occurred >45 Myr after the start of the Solar System, the $\epsilon\text{W}_{\text{BSE}}$ of zero can now be used to constrain conditions during accretion. Using $(\text{Hf}/\text{W})_{\text{BSE}} = 15$, a τ of ~ 13 Myr and a t_{GI} at 38 Myr are obtained with the model presented here. This small difference from $\tau \approx 11$ Myr (ref. 1) mainly reflects the different style of model that is smooth versus punctuated. If the $\epsilon\text{W}_{\text{BSSI}}$ was lower³, say -4.0 , the best estimate for t_{GI} would increase from 38 to 43 Myr. If $(\text{Hf}/\text{W})_{\text{BSE}}$ started much lower but then with the loss of volatiles during the giant impact W became more siderophile, the calculated t_{GI} would increase slightly from 38 to 41 Myr. All these modifications to conditions during accretion or the fundamental Hf–W parameters still result in calculated $t_{\text{GI}} < 45$ Myr.

In contrast, incomplete mixing and equilibration of impacting core material is a very effective way of producing Hf–W timescales that seem too short (Fig. 2b). The t_{GI} extends to >70 Myr with incomplete equilibration of Theia's core with the BSE. The timescale is even longer if partial equilibration includes earlier episodes (Fig. 2b). Therefore, the simplest explanation for faster W isotopic accretion rates is incomplete equilibration of W.

It is normally assumed that accreted W must approach equilibrium with the silicate Earth, otherwise the BSE would not be almost chondritic⁵. The small (0.2‰) excess ^{182}W that now has

Table 1 Calculated values for the Earth's accretion timescale based on lead isotopes

Estimate	Reference	$^{206}\text{Pb}/^{204}\text{Pb}$	$^{207}\text{Pb}/^{204}\text{Pb}$	μ_{BSE}	Model A		Model B		Model C	
					t_{GI}	τ	t_{GI}	τ	t_{GI}	τ
MKC-03	25	18.07	15.54	8.66	72	25	196	68	59	20
KC-99	26	18.27	15.60	8.91	90	31	253	87	73	25
KT-97	27	17.44	15.16	8.04	74	25	199	69	60	21
GG-91	28	18.11	15.617	8.67	54	19	147	51	44	15
L-91	29	17.92	15.47	8.50	65	23	177	61	53	18
K-89	30	17.822	15.445	8.38	54	19	145	50	44	15
AL-89	31	18.34	15.551	9.05	119	41	343	118	97	33
A-88	32	18.40	15.58	9.11	120	41	349	120	98	34
ZH-88	33	18.619	15.565	9.44	165	57	503	173	134	46
D-84	34	17.83	15.457	8.38	52	18	139	48	42	14
DZ-79	35	18.252	15.476	8.98	127	44	368	127	103	36

Calculated values for the τ and t_{GI} given in Myr as deduced from different estimates of the Pb isotopic composition of the BSE using the type of accretion model shown in Fig. 1. All calculations assume continuous core formation and total equilibration between accreted material and the BSE. All $^{238}\text{U}/^{204}\text{Pb}$ (μ) values are present-day equivalent values. The present-day $(^{206}\text{Pb}/^{204}\text{Pb})_{\text{BSE}}$ can be used to place accurate limits on the current μ_{BSE} because the age of the Earth introduces a trivial uncertainty and early changes in μ are insignificant for present-day $(^{206}\text{Pb}/^{204}\text{Pb})_{\text{BSE}}$. The single-stage Pb–Pb model age was used. Model A uses this μ_{BSE} to define the Pb depletion caused by core formation and to calculate the accretion timescales. Model A assumes no Pb loss from the Earth and $\mu_{\text{TOTE}} = 0.7$ (ref. 15). Model B assumes that Pb is depleted by a factor of only 2.5 by core formation as proposed in ref. 37. The μ_{TOTE} is then adjusted to yield the μ_{BSE} shown and is in the range 3 to 4. Model C is the same except that a $\mu_{\text{TOTE}} = 0.35$ is assumed until the giant impact when half the Earth's Pb is instantaneously lost. Such a model is dynamically difficult (see text). The μ of 0.35 merely represents the composition of all accreting material in this model. In every model $(^{206}\text{Pb}/^{204}\text{Pb})_{\text{BSSI}} = 9.307$, $(^{207}\text{Pb}/^{204}\text{Pb})_{\text{BSSI}} = 10.294$, $^{238}\text{U}/^{235}\text{U} = 137.88$, $\lambda^{238}\text{U} = 9.85 \times 10^{-10} \text{ yr}^{-1}$, $\lambda^{235}\text{U} = 1.55 \times 10^{-10} \text{ yr}^{-1}$.

been convincingly resolved^{1–3} may record some extreme event in the range of conditions, such as the giant impact. This is consistent with simulations of accretion and core formation^{6–10}. For illustrative purposes, if t_{GI} were 55 Myr the amount of equilibration of Theia's core with the BSE during the giant impact would have been 26%, assuming 100% before and after equilibration and constant $(\text{Hf}/\text{W})_{\text{BSE}}$ and $(\text{Hf}/\text{W})_{\text{BSI}}$ of 15, where BSI is the bulk silicate impactor (Fig. 3a). Note that the giant impact would increase $\epsilon\text{W}_{\text{BSE}}$. However, the initial W isotopic composition of the Moon indicates that Hf/W in planetary mantles was anything but constant, as explained next.

Mars-like protoplanets to the Earth–Moon system

The majority of giant impact simulations generate most of the Moon's mass from the silicate portion of Theia^{6,7}. Tungsten isotopes provide a powerful constraint on the Moon's precursor materials. If the Moon was produced from the BSI with $(\text{Hf}/\text{W})_{\text{BSI}} = (\text{Hf}/\text{W})_{\text{BSE}} = 15$, the initial ϵW of the lunar mantle would be >15 . This is because Mars-sized planets like Theia are thought to have completed their growth early, like Mars⁴⁷.

Instead, the W isotopic composition of the Moon is very similar

to the silicate Earth—slightly more radiogenic than the average Solar System as determined from carbonaceous chondrites². Even admixing 5% of metal from Theia's core and 45% of BSE does not reproduce the W isotopic composition of the Moon. This is already beyond the limit of the proportions demonstrated in dynamic simulations. It is unlikely that accretion of Theia was much longer than implied by theory and observations of Mars-sized objects. It is of course possible that a massive reduction in the $\epsilon\text{W}_{\text{BSI}}$ was produced by a major impact on Theia just before the giant impact. However, such an *ad hoc* low-probability explanation is unnecessary.

The simplest explanation is that $(\text{Hf}/\text{W})_{\text{BSI}}$ and possibly $(\text{Hf}/\text{W})_{\text{BSE}}$ too, were substantially less than in the Moon or silicate Earth today. If $(\text{Hf}/\text{W})_{\text{BSI}}$ and $(\text{Hf}/\text{W})_{\text{BSE}}$ were <3 times the chondritic value, an initial ϵW of zero to $+1$ for the early Moon could be generated. This time-integrated Hf/W is almost identical to the Hf/W in the silicate Mars^{21,47}. The explanation usually given is that the martian mantle was more oxidizing, leading to more lithophile behaviour of W. Independent evidence for greater oxidation of the mantle of Theia has been found in the higher Fe content of the lunar mantle⁴⁸, which again is like that inferred for Mars²¹.

The time-integrated Rb/Sr for Theia deduced from lunar samples is also identical to Mars and is significantly more volatile-rich than the Earth and Moon¹¹. Therefore, Rb losses and changes in oxidation state were both late (>10 Myr), well within the time realm of planetary growth via collisions. A simple explanation for the lunar data then is that volatile loss occurred from the material forming the Moon during the giant impact.

However, the similarity in oxygen isotopes provides evidence that the proto-Earth and Theia were made from a similar mix of materials⁴⁹. If this is the case, impact-driven losses of volatiles also contributed to the Rb-depletion and oxidation-state of the terrestrial mantle. There is indeed a relationship between the Hf/W and Rb/Sr of inner Solar System primitive mantles, as deduced so far (Fig. 4). Hafnium and W are both refractory elements and there is no obvious explanation for this correlation unless the oxidation of planetary mantles is somehow related to loss of moderately volatile elements. This could be late, as in the case of the Moon, or very early, as in the case of Vesta.

The direct loss of heavy lithophile elements against the gravita-

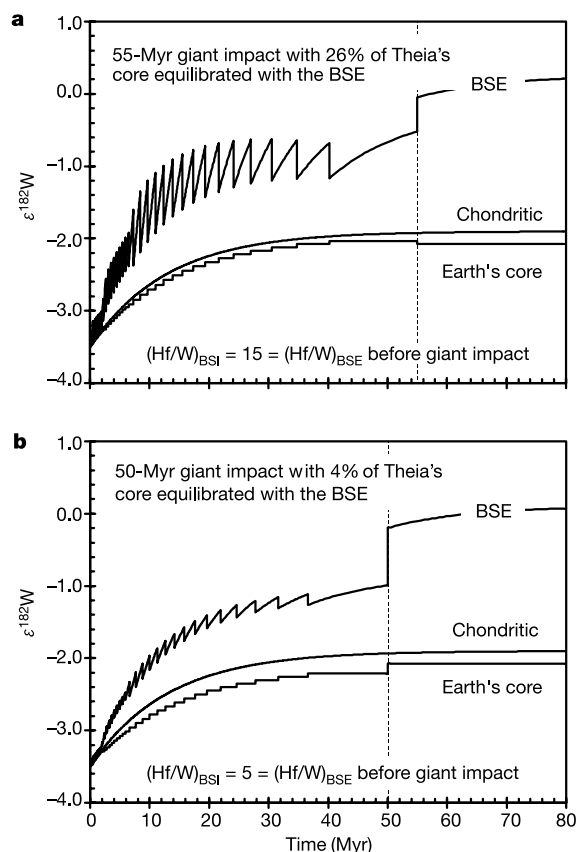


Figure 3 The effects of disequilibrium on tungsten isotopic changes in the BSE can be large. In both **a** and **b** the evolution of the total Earth (chondritic), core and BSE are calculated, assuming $\epsilon^{182}\text{W}_{\text{BSI}} = -3.5$. $\epsilon^{182}\text{W}$ are all expressed relative to the present day BSE. Note that in both cases the effect of the giant impact is to increase the $\epsilon^{182}\text{W}_{\text{BSE}}$. **a**, A present-day $\epsilon^{182}\text{W}_{\text{BSE}}$ of zero can be achieved if only 26% of Theia's core equilibrated with the BSE during a giant impact at 55 Myr. This model assumes that the Moon and Earth formed from protoplanets that had Hf/W in their silicate portions like that in the present Earth (15), which may be incorrect. **b**, In this model the Moon and Earth formed from protoplanets that had Hf/W in their silicate portions that was about a third of that in the present Earth, that is, more like Mars^{21,47}. A present-day $\epsilon^{182}\text{W}_{\text{BSE}}$ of zero can be achieved if only 4% of the W in Theia's core equilibrated with the BSE during a giant impact at 50 Myr.

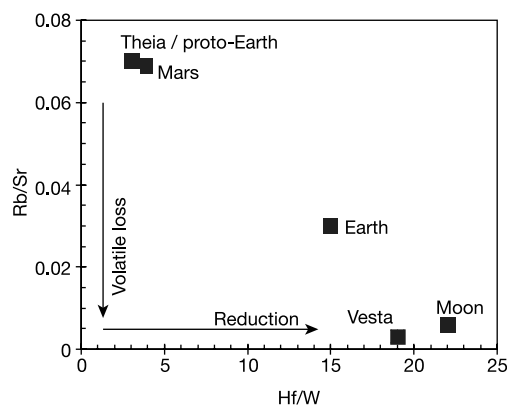


Figure 4 Inner Solar System primitive mantles. The Hf/W appears to be negatively correlated with Rb/Sr in the primitive mantles of inner Solar System planetesimals and planets. A possible explanation for this is that the loss of moderately volatile elements was linked to loss of other volatiles during planetary collisions, such that the mantles changed from more oxidizing to more reducing. The Moon is an extreme example of this. That Theia is so like Mars in these respects provides evidence that such volatile-rich objects may have dominated the entire inner Solar System during the early stages of planetary accretion. The values for Theia are time-averaged (ref. 11 and this study).

tional force of the Earth is unlikely^{11,48}. Losses from smaller accreting protoplanets are more readily accommodated. Therefore, the BSE composition shown in Fig. 4 may reflect admixing of planets and planetesimals that lost different amounts of volatiles at an earlier stage. However, it may also be that moderately volatile elements like Rb were partially atmophile because of high surface temperatures in hot early planetary environments^{18,39}. Early proto-atmospheres were lost from the Earth, as indicated by Xe data³⁹, and very probably also from constituent protoplanets—possibly on more than one occasion. If moderately volatile incompatible elements were partially circulated through hot atmospheres¹⁸ they may well have been lost during blow-off events⁴⁰. A partially atmophile behaviour might also help to explain the greater isotopic equilibration of moderately volatile Pb versus refractory W during accretion.

The potential effects of changes in oxidation on the W isotopic evolution of the Earth could be significant, whether caused by changes in the composition of accreted material or blow-off (Fig. 3b). For example, growth from protoplanets with Hf/W = 5 until a giant impact at 50 Myr leading to volatile loss and a corresponding increase in (Hf/W)_{BSE} to the value of 15 found today achieves a $\epsilon^{182}\text{W}_{\text{BSE}}$ of zero with only 4% equilibration between silicate and newly accreted metal during the giant impact. The abundances of highly siderophile elements in the upper mantle⁵⁰ could then relate to this event, plus the superimposed effects of the late veneer^{23,48}. Furthermore, much of the variability in moderately volatile elements and oxidation in the inner Solar System may be the cumulative effect of collisions between early-formed planetesimals that were, in general, more Mars-like. □

Received 29 August; accepted 9 December 2003; doi:10.1038/nature02275.

- Yin, Q. Z. *et al.* A short timescale for terrestrial planet formation from Hf–W chronometry of meteorites. *Nature* **418**, 949–952 (2002).
- Kleine, T., Münker, C., Mezger, K. & Palme, H. Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf–W chronometry. *Nature* **418**, 952–955 (2002).
- Schönberg, R., Kamber, B. S., Collerson, K. D. & Eugster, O. New W isotope evidence for rapid terrestrial accretion and very early core formation. *Geochim. Cosmochim. Acta* **66**, 3151–3160 (2002).
- Jacobsen, S. B. & Harper, C. L. Jr in *Earth Processes: Reading the Isotope Code* (eds Basu, A. & Hart, S.) 47–74 (AGU, Washington DC, 1996).
- Halliday, A. N. Terrestrial accretion rates and the origin of the Moon. *Earth Planet. Sci. Lett.* **176**, 17–30 (2000).
- Cameron, A. G. W. & Benz, W. Origin of the Moon and the single impact hypothesis IV. *Icarus* **92**, 204–216 (1991).
- Canup, R. M. & Asphaug, E. Origin of the moon in a giant impact near the end of the Earth's formation. *Nature* **412**, 708–712 (2001).
- Yoshino, T., Walter, M. J. & Katsura, T. Core formation in planetesimals triggered by permeable flow. *Nature* **422**, 154–157 (2003).
- Rubie, D. C., Melosh, H. J., Reid, J. E., Liebske, C. & Righter, K. Mechanisms of metal-silicate equilibration in the terrestrial magma ocean. *Earth Planet. Sci. Lett.* **205**, 239–255 (2003).
- Stevenson, D. J. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 231–249 (Oxford Univ. Press, Oxford, 1990).
- Halliday, A. N. & Porcelli, D. In search of lost planets – the paleocosmochemistry of the inner solar system. *Earth Planet. Sci. Lett.* **192**, 545–559 (2001).
- Humayun, M. & Clayton, R. N. Potassium isotope cosmochemistry: Genetic implications of volatile element depletion. *Geochim. Cosmochim. Acta* **59**, 2131–2151 (1995).
- Wetherill, G. W. in *Origin of the Moon* (eds Hartmann, W. K., Phillips, R. J. & Taylor, G. J.) 519–555 (Lunar Planetary Institute, Houston, 1986).
- Taylor, S. R. & Norman, M. D. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 29–43 (Oxford Univ. Press, Oxford, 1990).
- Allègre, C. J., Manhès, G. & Göpel, C. The age of the Earth. *Geochim. Cosmochim. Acta* **59**, 1445–1456 (1995).
- Lee, D.-C. & Halliday, A. N. Hafnium–tungsten chronometry and the timing of terrestrial core formation. *Nature* **378**, 771–774 (1995).
- Galer, S. J. G. & Goldstein, S. L. in *Earth Processes: Reading the Isotope Code* (eds Basu, A. & Hart, S.) 75–98 (AGU, Washington DC, 1996).
- Hayashi, C., Nakazawa, K. & Nakagawa, Y. in *Protostars and Planets II* (eds Black, D. C. & Matthews, M. S.) 1100–1153 (Univ. Arizona Press, Tucson, 1985).
- Sasaki, S. & Nakazawa, K. Metal-silicate fractionation in the growing Earth: energy source for the terrestrial magma ocean. *J. Geophys. Res.* **91**, B9231–B9238 (1986).
- Carlson, R. W. & Lugmair, G. W. in *Origin of the Earth and Moon* (eds Canup, R. & Righter, K.) 25–44 (Univ. Arizona Press, Tucson, 2000).
- Newsom, H. E. in *Global Earth Physics, A Handbook of Physical Constants* (ed. Ahrens, T. J.) 159–189 (AGU Reference Shelf 1, American Geophysical Union, Washington DC, 1995).
- Lissauer, J. J. Time-scales for planetary accretion and the structure of the protoplanetary disk. *Icarus* **69**, 249–265 (1987).
- Newsom, H. E. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 273–288 (Oxford Univ. Press, Oxford, 1990).
- Patterson, C. C. Age of meteorites and the Earth. *Geochim. Cosmochim. Acta* **10**, 230–237 (1956).
- Murphy, D. T., Kamber, B. S. & Collerson, K. D. A refined solution to the first terrestrial Pb-isotope paradox. *J. Petrol.* **44**, 39–53 (2003).
- Kamber, B. S. & Collerson, K. D. Origin of ocean-island basalts: a new model based on lead and helium isotope systematics. *J. Geophys. Res.* **104**, 25479–25491 (1999).
- Kramers, J. D. & Tolstikhin, I. N. Two terrestrial lead isotope paradoxes, forward transport modeling, core formation and the history of the continental crust. *Chem. Geol.* **139**, 75–110 (1997).
- Galer, S. J. G. & Goldstein, S. L. Depleted mantle Pb isotopic evolution using conformable ore leads. *Terra Abstr.* **3**, 485–486 (1991).
- Liew, T. C., Milisenda, C. C. & Hofmann, A. W. Isotopic contrasts, chronology of element transfers and high-grade metamorphism: the Sri Lanka Highland granulites, and the Lewisian (Scotland) and Nuk (S.W. Greenland) gneisses. *Geol. Rundsch.* **80**, 279–288 (1991).
- Kwon, S.-T., Tilton, G. R. & Grünfelder, M. H. in *Carbonatites—Genesis and Evolution* (ed. Bell, K.) 360–387 (Unwin-Hyman, London, 1989).
- Allègre, C. J. & Lewin, E. Chemical structure and history of the Earth: Evidence from global non-linear inversion of isotopic data in a three box model. *Earth Planet. Sci. Lett.* **96**, 61–88 (1989).
- Allègre, C. J., Lewin, E. & Dupré, B. A coherent crust-mantle model for the uranium-thorium-lead isotopic system. *Chem. Geol.* **70**, 211–234 (1988).
- Zartman, R. E. & Haines, S. M. The plumbotectonic model for Pb isotopic systematics among major terrestrial reservoirs—a case for bi-directional transport. *Geochim. Cosmochim. Acta* **52**, 1327–1339 (1988).
- Davies, G. F. Geophysical and isotopic constraints on mantle convection: An interim synthesis. *J. Geophys. Res.* **89**, 6017–6040 (1984).
- Doe, B. R. & Zartman, R. E. in *Geochemistry of Hydrothermal Ore Deposits* (ed. Barnes, H. L.) 22–70 (Wiley, New York, 1979).
- Davies, G. F. Geophysically constrained mass flows and the ⁴⁰Ar budget: a degassed lower mantle? *Earth Planet. Sci. Lett.* **166**, 149–162 (1999).
- Azbel, I. Y., Tolstikhin, I. N., Kramers, J. D., Pechernikova, G. V. & Vityazev, A. V. Core growth and siderophile element depletion of the mantle during homogeneous Earth accretion. *Geochim. Cosmochim. Acta* **57**, 2889–2898 (1993).
- Ozima, M. & Podosek, F. A. Formation age of Earth from ¹²⁹I/¹²⁷I and ²⁴⁴Pu/²³⁸U systematics and the missing Xe. *J. Geophys. Res.* **104**, 25493–25499 (1999).
- Porcelli, D., Cassen, P. & Woolum, D. Deep Earth rare gases: Initial inventories, capture from the solar nebula and losses during Moon formation. *Earth Planet. Sci. Lett.* **193**, 237–251 (2001).
- Benz, W. & Cameron, A. G. W. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 61–67 (Oxford Univ. Press, Oxford, 1990).
- Tera, F., Papanastassiou, D. A. & Wasserburg, G. J. A lunar cataclysm at ~3.95 AE and the structure of the lunar crust. *Lunar Planet. Sci.* **IV**, 723–725 (1973).
- Hanan, B. B. & Tilton, G. R. 60025: Relict of primitive lunar crust? *Earth Planet. Sci. Lett.* **84**, 15–21 (1987).
- Carlson, R. W. & Lugmair, G. W. The age of ferroan anorthosite 60025: oldest crust on a young Moon? *Earth Planet. Sci. Lett.* **90**, 119–130 (1988).
- Lee, D.-C., Halliday, A. N., Snyder, G. A. & Taylor, L. A. Age and origin of the Moon. *Science* **278**, 1098–1103 (1997).
- Leya, I., Wieler, R. & Halliday, A. N. Cosmic-ray production of tungsten isotopes in lunar samples and meteorites and its implications for Hf–W cosmochemistry. *Earth Planet. Sci. Lett.* **175**, 1–12 (2000).
- Lee, D.-C., Halliday, A. N., Leya, I., Wieler, R. & Wiechert, U. Cosmogenic tungsten and the origin and earliest differentiation of the Moon. *Earth Planet. Sci. Lett.* **198**, 267–274 (2002).
- Lee, D.-C. & Halliday, A. N. Core formation on Mars and differentiated asteroids. *Nature* **388**, 854–857 (1997).
- Jones, J. H. & Palme, H. in *Origin of the Earth and Moon* (eds Canup, R. & Righter, K.) 197–216 (Univ. Arizona Press, 2000).
- Wiechert, U. *et al.* Oxygen isotopes and the Moon-forming giant impact. *Science* **294**, 345–348 (2001).
- Righter, K. & Drake, M. J. Effect of water on metal-silicate partitioning of siderophile elements: a high pressure and temperature terrestrial magma ocean and core formation. *Earth Planet. Sci. Lett.* **171**, 383–399 (1999).

Acknowledgements This paper benefited from discussions with, and comments from, T. Ahrens, W. Benz, R. Canup, R. Carlson, M. Drake, T. Grove, M. Humayun, T. Kleine, K. Mezger, C. Münker, H. Palme, D. Porcelli, M. Schönbachler, D. Stevenson, M. Walter, R. Wieler and H. Williams and was supported by ETH and Swiss National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.N.H. (halliday@erdw.ethz.ch).

The conserved kinetochore protein shugoshin protects centromeric cohesion during meiosis

Tomoya S. Kitajima¹, Shigehiro A. Kawashima¹ & Yoshinori Watanabe^{1,2}

¹Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, and ²SORST, Japan Science and Technology Agency, Hongo, Tokyo 113-0033, Japan

Meiosis comprises a pair of specialized nuclear divisions that produce haploid germ cells. To accomplish this, sister chromatids must segregate together during the first meiotic division (meiosis I), which requires that sister chromatid cohesion persists at centromeres. The factors that protect centromeric cohesion during meiosis I have remained elusive. Here we identify Sgo1 (shugoshin), a protector of the centromeric cohesin Rec8 in fission yeast. We also identify a homologue of Sgo1 in budding yeast. We provide evidence that shugoshin is widely conserved among eukaryotes. Moreover, we identify Sgo2, a paralogue of shugoshin in fission yeast, which is required for faithful mitotic chromosome segregation. Localization of Sgo1 and Sgo2 at centromeres requires the kinase Bub1, identifying shugoshin as a crucial target for the kinetochore function of Bub1. These findings provide insights into the evolution of meiosis and kinetochore regulation during mitosis and meiosis.

In eukaryotes, sister chromatid cohesion is established during S phase and is maintained throughout G2 until the M phase. During mitosis, this cohesion is destroyed along the entire length of the chromosome, allowing sister chromatids to segregate to opposite sides of the cell (equational division), and ensuring that each daughter cell receives one copy of each chromosome. In contrast, meiosis consists of two rounds of chromosome segregation following a single round of DNA replication, leading to the formation of four haploid gametes from a diploid germ cell. During meiosis I, homologous chromosomes (homologues) pair up in order to recombine, forming chiasmata in which one sister chromatid from one homologue is covalently attached to a sister chromatid from the other homologue. Hence, for homologues to segregate at meiosis I, sister chromatid cohesion must be released along the chromosome arms to resolve chiasmata. However, sister chromatid cohesion is retained at the centromeres until meiosis II, when sister chromatids segregate as they do in mitosis, using the residual centromeric cohesion. Thus, meiotic divisions require sister chromatid cohesion to be released in two steps, yet the molecular basis for protection of centromeric cohesion only during meiosis I and only at the centromeres has remained unknown¹.

There are clues to the molecular nature of sister chromatid cohesion and the mechanism by which it is released at the onset of anaphase^{1–5}. In various eukaryotes, sister chromatid cohesion depends on a multisubunit cohesin complex including Scc1 (Rad21 in the fission yeast *Schizosaccharomyces pombe*). Anaphase-promoting complex (APC)-dependent degradation of the securin Pds1 (Cut2 in *S. pombe*) allows release of the Esp1 (Cut1 in *S. pombe*) endopeptidase (separase), which in turn cleaves Scc1, releasing sister chromatid cohesion. During meiosis, the cohesin subunit Scc1 is replaced by a meiotic counterpart, Rec8 (refs 6–10). As Rec8 complexes reside only at centromeres after meiosis I and depletion of Rec8 disrupts centromeric cohesion, its presence at centromeres has been thought to confer the persistence of cohesion throughout meiosis I (ref. 11). Several lines of evidence^{12,13} suggest that Rec8 along chromosome arms is cleaved by separase at anaphase I, whereas centromeric Rec8 is specifically protected until metaphase II. Budding yeast Spo13 has been implicated in the protection of centromeric Rec8 (refs 14, 15), but Spo13 is not centromeric and may function indirectly. *Drosophila* MEI-S332, a protein that resides

at pericentromeric regions¹⁶ and is required for the persistence of centromeric cohesion during meiosis I (ref. 17), has features of a candidate protector of meiotic centromeric cohesion, although the details of such protection have so far not been revealed⁴. Despite the completion of genome sequencing projects in several organisms, no homologues of Spo13 or MEI-S332 have emerged, preventing the formulation of a generalized view of protection. Concurrently, studies in fission yeast¹⁸ have illuminated the importance of pericentromeric heterochromatin for recruiting centromeric Rec8 complexes and ensuring centromeric cohesion during meiosis I. However, pericentromeric heterochromatin cannot alone confer the specific protection of Rec8 at meiosis I compared with meiosis II. We now identify a meiosis-specific protein, Sgo1 (shugoshin, Japanese for ‘guardian spirit’), that protects centromeric Rec8 from degradation during meiosis I.

Identification of Sgo1 in fission yeast

The replacement of the mitotic cohesin Rad21 with the meiotic version Rec8 is a prerequisite for protecting centromeric sister chromatid cohesion through anaphase of meiosis I (refs 19, 20). However, when we expressed Rec8 ectopically during mitosis, Rec8 localized largely at centromeres but disappeared at anaphase, with sister chromatids segregating to opposite sides of the cell (Fig. 1c, d). Moreover, the ectopic expression of non-cleavable Rec8 during mitosis (note that Rec8 is cleaved by separase Cut1 during meiosis¹³) resulted in an inability of sister chromatids to separate (Supplementary Fig. 1). Thus, in contrast to the situation during meiosis I, centromeric Rec8 is cleaved by separase during mitosis, resulting in sister chromatid separation. These observations prompted us to postulate a meiosis-I-specific centromeric protector of Rec8. To identify this factor, we searched for a gene that yields toxicity during mitotic growth only when co-expressed with Rec8. This screen identified a novel gene called *sgo1*⁺ (open reading frame (ORF) SPBP35G2.03C). The hindrance of growth by Sgo1 was indeed dependent on Rec8, as Sgo1 had little effect on growth when co-expressed with Rad21 (Fig. 1a). Co-expression of *rec8*⁺ and *sgo1*⁺ frequently led to blocked nuclear division, as centromere-associated green fluorescent protein markers (*cen2-GFP*)²¹ frequently segregated to the same side of a septated cell (Fig. 1b, c). To test the possibility that Sgo1 protects Rec8 from degradation at anaphase,

we examined the localization of Rec8 in the context of Sgo1 expression. We found that the Rec8 localization at centromeres persisted through anaphase only when Sgo1 was co-expressed (Fig. 1d). As Sgo1 is expressed exclusively in meiosis (DNA microarray data²²; see below), the foregoing results allowed us to postulate that Sgo1 is a protector of Rec8 during meiosis.

Sgo1 protects centromeric cohesion at meiosis I

To examine whether Sgo1 is indeed required for the protection of Rec8 during meiosis, we deleted the entire ORF encoding *sgo1*⁺ and examined the phenotype. *sgo1*-depleted cells (*sgo1*Δ) were viable and showed normal vegetative growth. To examine meiotic chromosome segregation, we marked centromere-linked sequences with GFP (*cen2*-GFP) on only one of the two homologues in a zygote and monitored the segregation of the GFP dots during meiosis. In normal meiosis, monopolar attachment of sister kinetochores to the spindle is established in metaphase I; therefore, sisters move together to the same side of the zygote (reductional division) in the following anaphase I (Fig. 2a). Thanks to the centromeric cohesion preserved throughout anaphase I, bipolar attachment is

secured at meiosis II, thus leading to faithful disjunction (Fig. 2a). We found that meiosis I appeared normal in *sgo1*Δ cells, as sister chromatid pairs generally moved together to the same side of each zygote (Fig. 2b). Hence, monopolar attachment was intact in this mutant. Moreover, by marking *cen2*-GFP on both chromosomes, we determined that homologues underwent faithful disjunction at meiosis I (data not shown). At meiosis II in *sgo1*Δ cells, however, sister chromatids failed to segregate properly, undergoing non-disjunction in approximately 50% of cells (Fig. 2b). This value is consistent with random chromosome segregation at meiosis II.

To examine centromeric cohesion, we monitored *cen2*-GFP marked on both homologues in zygotes arrested before meiosis II, the stage at which centromeric cohesion is normally preserved in *sgo1*⁺ cells. We found that *sgo1*Δ cells frequently displayed precocious centromeric dissociation as split *cen2*-GFP signals prevailed in the dyad nuclei (Fig. 2c). This result may explain why the second meiotic division is random in *sgo1*Δ cells, because cohesion is required for sister kinetochores to be properly recognized by spindle microtubules extending from opposite poles. Next we examined whether protection of Rec8 at centromeres is dependent on Sgo1 by monitoring Rec8-GFP at late anaphase I and prometaphase II. Notably, although Rec8 signals were centromeric in wild-type cells, the Rec8 signals had largely disappeared from the centromeres at these stages in *sgo1*Δ cells (Fig. 2d). All phenotypes of *sgo1*Δ cells are reminiscent of heterochromatin-deficient *S. pombe*, in which Rec8 localization to the pericentromeric regions is decreased and centromeric cohesion is lost during meiosis I, leading to random division at meiosis II (ref. 18). We then examined chromatin binding by Rec8 in cells arrested before meiosis I using a chromatin immunoprecipitation (ChIP) assay. In marked contrast to heterochromatin-deficient cells, Rec8 localization was intact in *sgo1*Δ cells at the pericentromeric regions as well as all other regions tested (Fig. 2e). As monopolar attachment requires centromeric Rec8 (refs 20, 23), the proper location of Rec8 before meiosis I is consistent with the fact that monopolar attachment is intact in *sgo1*Δ cells. More importantly, these results indicate that the loss of centromeric Rec8 after meiosis I is brought about not by an initial defect in Rec8 localization to centromeres but rather by a defect in the preservation of centromeric Rec8 during anaphase I. Previous results suggested that the Cut1 separase becomes active at the onset of anaphase I and cleaves most chromosomal Rec8, leaving only centromeric Rec8 intact¹³. Our current results advocate that Sgo1 has an essential role in protecting centromeric cohesion at anaphase I by safeguarding cohesin Rec8 from separase cleavage.

Sgo1 localizes at centromeres during meiosis I

To detect the Sgo1 protein, we raised antibodies specific for Sgo1. Western blotting indicated that Sgo1 is expressed only around meiosis I (Fig. 3a). Immunofluorescence microscopy on cells at various stages of meiosis revealed that Sgo1 appears at late prophase of meiosis I and is fully localized as several punctate dots until metaphase I (Fig. 3b). These dots localize closely with the Mis6 kinetochore protein²⁴, indicating that Sgo1 is associated with centromeric regions (Fig. 3c). At the onset of anaphase I, Sgo1 signals decrease markedly. We found that Sgo1 remains undegraded at centromeres in APC-depleted cells arrested at metaphase I, but undergoes normal degradation in separase-defective cells (Supplementary Fig. 2), suggesting that Sgo1 degradation at anaphase I is regulated more directly by the APC rather than through separase. Although we detected residual Sgo1 signals at the centromeres in early anaphase I, they disappeared completely by the end of anaphase I (Fig. 3b). This suggests that a substantial amount of Sgo1 is required at the onset of anaphase I when separase is fully activated. As anaphase I progresses, however, smaller and smaller amounts of Sgo1 would be required. This idea is tenable if the separase activity is quickly downregulated or prevented from accessing chromosomes during anaphase I. Sgo1 does not reappear

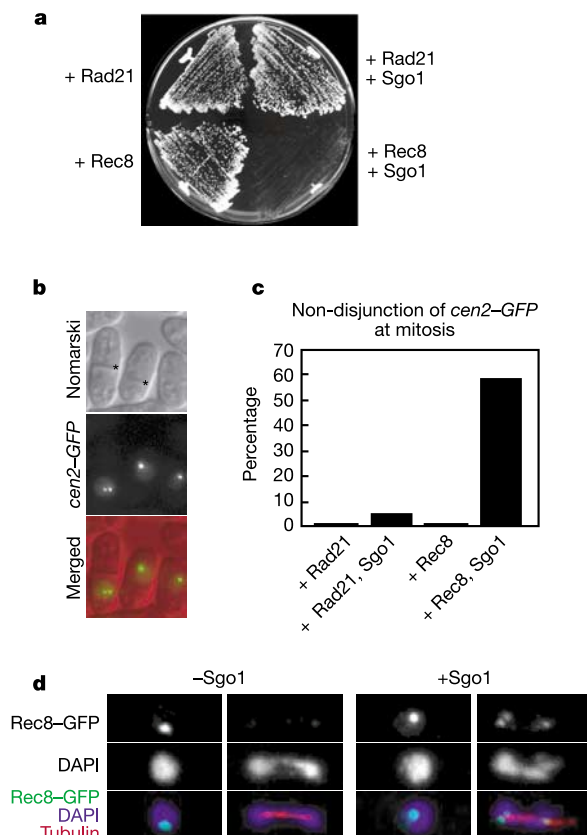


Figure 1 Co-expression of Sgo1 and Rec8 causes failure of sister chromatid separation during mitosis. **a**, The haploid *cen2*-GFP strains expressing the indicated genes by exogenous promoters (a constitutive promoter *Padh1* for *rad21*⁺ or *rec8*⁺, and a thiamine-repressible promoter *Pnmt1* for *sgo1*⁺) were streaked on a thiamine-depleted plate. **b**, Examples of *Padh1*-*rec8*⁺ *Pnmt1*-*sgo1*⁺ cells cultured at 30 °C for 15 h after thiamine depletion. Note the non-disjunction of *cen2*-GFP in the septated cells (asterisk). **c**, The frequency of non-disjunction was counted among septated cells (*n* > 100). **d**, *Padh1*-*rec8*⁺-GFP strains with or without *Pnmt1*-*sgo1*⁺ were cultured as in **b**, fixed with formaldehyde and stained with DAPI and antibodies against GFP and tubulin. Examples of each strain at interphase (left cell) and anaphase (right cell) are shown. Note that the Rec8 signal persists during anaphase mostly in *sgo1*⁺-expressing cells (84%, *n* = 129) and only slightly in non-expressing cells (9%, *n* = 129).

during meiosis II (Fig. 3b), consistent with the idea that Sgo1 is only required for the protection of Rec8 during meiosis I.

We have suggested previously that Rec8 localization at pericentromeric regions is especially important for the persistence of centromeric cohesion throughout meiosis I (ref. 18). If Sgo1 is a centromeric protector of Rec8, then it might be expected to localize there as well. To test this possibility, we delineated Sgo1 localization more precisely using a ChIP assay. Sgo1 does indeed associate with pericentromeric heterochromatin regions rather than central core regions along the centromere sequences (Fig. 3d). As immunoprecipitation experiments indicated that Sgo1 interacts with Rec8 complexes *in vivo* (Fig. 3f), the protection is probably carried out through close interaction. Together, our results indicate that Sgo1 resides at pericentromeric regions and operates to protect Rec8 from cleavage by separase at anaphase I (Fig. 3g).

Sgo2 is a mitotic Sgo1 paralogue in fission yeast

By means of a conventional BLAST search of genome databases, we identified Sgo1-like proteins from *Saccharomyces cerevisiae* and *Neurospora crassa*, suggesting that Sgo1 is a conserved protein (see below). In the same search, we identified an *S. pombe* Sgo1 paralogue, which we call Sgo2 (ORF SPAC15A10.15). We disrupted the *sgo2*⁺ gene and found that *sgo2*Δ cells are viable but show sensitivity to the spindle-destabilizing drug thiabendazole (TBZ) (Fig. 4a). Consistently, *sgo2*Δ cells display an increased incidence of chromosome mis-segregation at mitosis (Fig. 4c). These mitotic phenotypes are remarkable, as *sgo1*Δ cells never show such a defect (Fig. 4a). To investigate its cellular distribution, the endogenous *sgo2*⁺ gene was tagged with GFP. In proliferating cells, Sgo2–GFP was observed as two or three dots within the nucleus (Fig. 4e); however, it localized closely with the centromere protein Mis6 at

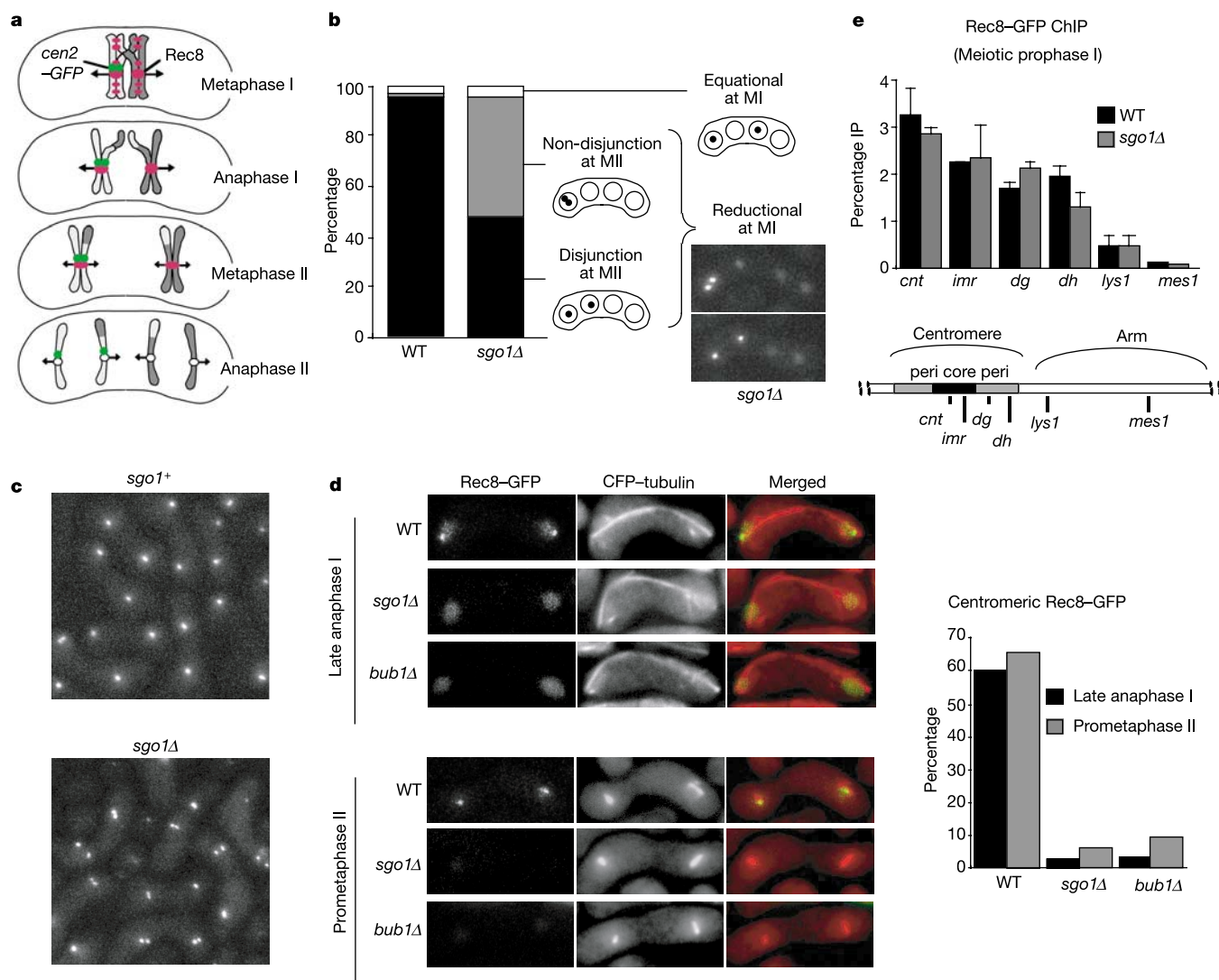


Figure 2 Sgo1 is required to protect Rec8 and thereby cohesion at centromeres during anaphase of meiosis I. **a**, Schematic drawing of the behaviour of homologous chromosomes (white and grey) in normal meiosis I and II. Expression of *cen2-GFP* (green oval) is marked on one of the homologous chromosomes and the location of the cohesin Rec8 (red oval) is indicated. **b**, One of the homologues marked with *cen2-GFP* was monitored for segregation during meiosis in wild-type (WT) and *sgo1*Δ cells ($n > 170$). Examples of *sgo1*Δ cells are shown (bottom right panel). **c**, Cells of both homologues marked with *cen2-GFP* were arrested before meiosis II by introducing the *mes1*Δ

mutation, and were examined for *cen2-GFP* dots. **d**, The Rec8–GFP signal was monitored at late anaphase I ($n > 30$) and at prometaphase II ($n > 100$) in the indicated cells, and the frequency of the cells displaying centromeric Rec8–GFP was counted. The spindles were visualized by expressing cyan fluorescent protein (CFP)–Atb2 (α 2-tubulin)⁴³. **e**, A ChIP assay with anti-GFP antibodies was used to measure Rec8–GFP levels throughout the indicated chromosome sites in the arrested cells before meiosis I (*mei4*Δ arrest). The bottom panel shows a schematic representation of *S. pombe* chromosome I as well as the primers used (*cnt*, *imr*, *dg*, *dh*, *lys1*, *mes1*).

metaphase and disappeared during anaphase (Fig. 4d, e). ChIP assays showed that Sgo2 chromatin association was detectable only in synchronous populations of mitotic cells, in which Sgo2 localized to the pericentromeric regions (Fig. 4f). Reinforcing this localization, *sgo2* deletion confers a marked defect in chromosome segregation if combined with the heterochromatin-deficient *swi6Δ* mutation, which by itself slightly impairs kinetochore function²⁵ (Fig. 4b, c). These results suggest that Sgo2 cooperates with pericentromeric heterochromatin factors to ensure chromosome segregation at mitosis. Moreover, we found that Sgo2 persists throughout meiosis (Supplementary Fig. 4) and that *sgo2Δ* cells have a modest increase in non-disjunction of homologues at meiosis I (about 15%), suggesting that Sgo2 is also important for promoting proper meiosis I. Notably, the role of Sgo2 in meiosis does not overlap with that of Sgo1, as *sgo1Δ* neither yields an apparent defect at meiosis I (Fig. 2b) nor enhances the defect of *sgo2Δ* (data not shown).

Shugoshin location controlled by Bub1

A conserved centromere-associated kinase, Bub1, is thought to have a function in protecting Rec8 during meiosis, as centromeric Rec8 cannot be detected after meiosis I in fission yeast *bub1* mutants²⁶ (Fig. 2d). Although *bub1* mutation has pleiotropic effects in meiotic chromosome segregation²⁶, we thought that Sgo1 function might be targeted by Bub1 activity. To address this issue, we examined Sgo1-GFP signals in *bub1Δ* cells undergoing meiosis. Notably, *bub1Δ* cells were almost completely devoid of punctate centromeric Sgo1-GFP signals, showing instead a diffuse fluorescence within the

nucleus (Fig. 3e). Identical results were obtained using the Bub1(K762R) point mutation (not shown), which abolishes the kinase activity²⁷. As substantial levels of Sgo1 protein were detected in meiotic *bub1Δ* cells by western blot analysis (Fig. 3e), Bub1 primarily regulates localization rather than stability of the Sgo1 protein. Therefore, the observed defects in centromeric protection in *bub1Δ* cells might be explained by an impaired location of Sgo1. Moreover, these results suggest that Bub1 might regulate the timing of Sgo1 loading to and the disappearance from centromeres, as Bub1 localizes at centromeres during a similar period in meiosis I (ref. 26).

In parallel experiments, we found that mitotic Sgo2 localization at centromeres was similarly disrupted in *bub1* mutants (Fig. 4d), whereas protein levels of Sgo2 were unchanged (Supplementary Fig. 3c). It has been suggested that loss of Bub1 function leads to a weakness in kinetochore function²⁸. In this regard, the Bub1(K762R) mutation shows a synthetic defect in growth with *swi6Δ*, a mutation that also impairs slightly kinetochore function through its role in pericentric heterochromatin formation. We found that *sgo2Δ* similarly showed a synthetic defect in growth with *swi6Δ* (Fig. 4b), exhibiting severe mis-segregation of chromosomes at mitosis (Fig. 4c). As the *sgo2Δ bub1Δ* double mutant showed no cumulative defects in growth or in TBZ sensitivity (Fig. 4a), these genetic analyses confirm that Sgo2 and Bub1 function in tandem to ensure chromosome segregation in mitosis. Taken together, these results reveal that Sgo1 and Sgo2 localization at centromeres is a crucial function of Bub1 kinase in meiosis and mitosis, respectively.

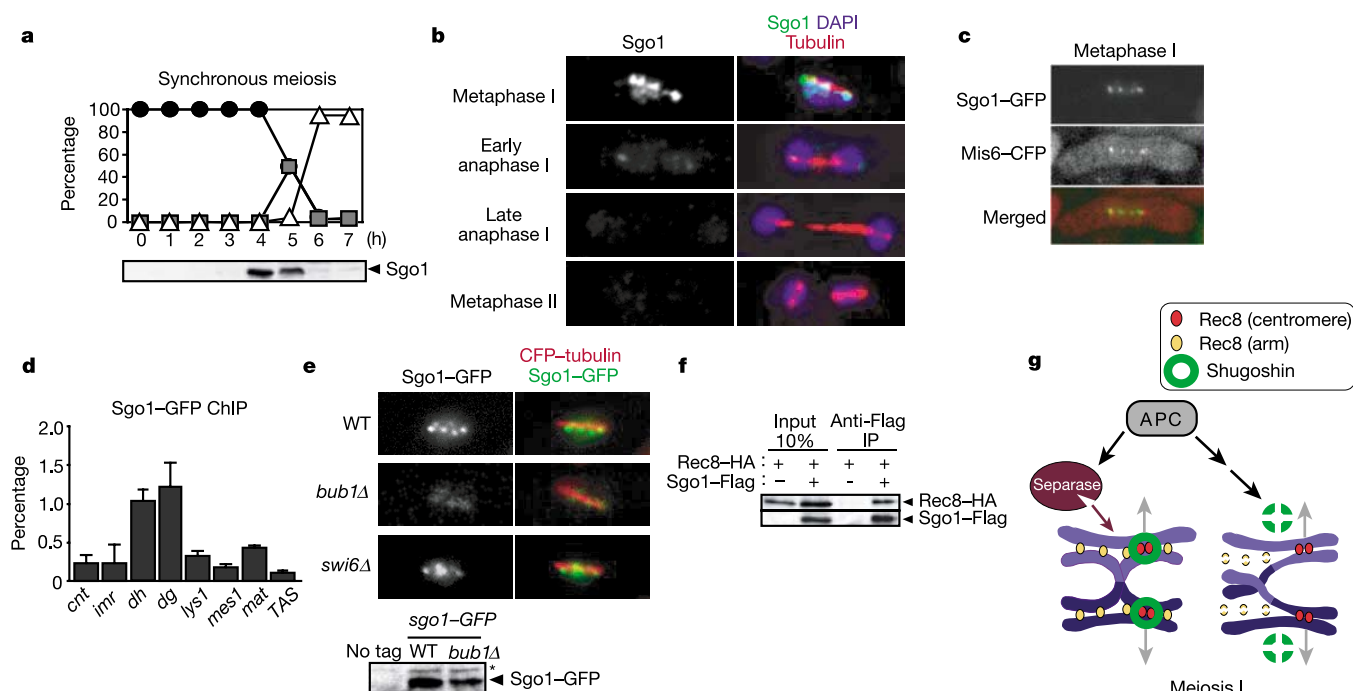


Figure 3 Sgo1 localizes at pericentromeric regions during meiosis I. **a**, Synchronous meiosis of diploid *pat1-114/pat1-114* cells¹³ was sampled. Meiotic nuclear division was monitored by DAPI staining, and the protein level of Sgo1 was detected by western blotting using anti-Sgo1 antibodies. Circle, one nucleus; square, two nuclei; triangle, 3–4 nuclei. **b**, Sgo1 (green) was counterstained with tubulin (red) and DAPI (blue) in the meiotic cell at the indicated stages. **c**, An *sgo1*⁺-GFP cell co-expressing *mis6*⁺-CFP was examined under fluorescence microscopy. Sgo1-GFP (green) and Mis6-CFP (red) are merged. **d**, A ChIP assay with anti-GFP antibodies was used to measure Sgo1-GFP levels throughout the indicated chromosome sites in cells arrested at metaphase I. We used the same primers as for Fig. 2e together with additional primers at *mat* (heterochromatin

region at the mating-type locus) and *TAS* (telomere-associated sequence). **e**, Sgo1-GFP (green) was detected at metaphase I in the indicated cells expressing CFP-Atb2 to visualize spindles (red). Western blot analysis of Sgo1-GFP was carried out on the indicated strains arrested at metaphase I. The asterisk indicates a nonspecific band. **f**, Rec8-HA was expressed with or without Sgo1-Flag in proliferating cells and the extracts were immunoprecipitated with anti-Flag antibody. **g**, A model for the action of shugoshin in meiosis. Shugoshin localizes at centromeres in meiosis I depending on Bub1 kinase, and it protects centromeric Rec8 complexes from cleavage by separase at the onset of anaphase I, thereby preserving centromeric cohesion until meiosis II. Shugoshin is degraded depending on APC during anaphase I.

Characterization of a Sgo1 homologue in budding yeast

We identified a single Sgo1 homologue in *S. cerevisiae*, ScSgo1 (ORF YOR073W), which has so far not been analysed. We examined the cellular localization of ScSgo1 by tagging endogenous ScSGO1 with GFP, and found that the pattern of ScSgo1 localization closely resembles that of *S. pombe* Sgo2 in mitosis and Sgo1 in meiosis (Supplementary Fig. 5). To examine the function of ScSgo1, we disrupted the ScSGO1 gene. *Scsgo1Δ* cells were viable but grew slowly and showed sensitivity to the spindle-destabilizing drug benomyl (Fig. 5a), suggesting that kinetochore function might be impaired. We used a colony-sectoring assay to compare the rates of chromosome loss in *Scsgo1Δ* cells with those in wild-type cells. Whereas less than 2% of wild-type colonies contained red sectors (which indicate chromosome loss), approximately 40% of the *Scsgo1Δ* colonies contained such sectors (Fig. 5b). We conclude that ScSgo1 has a crucial role at kinetochores for ensuring mitotic chromosome segregation. *Scsgo1Δ* cells showed significant defects in the initiation of meiosis, as many cells arrested with a single nucleus in the meiotic condition. Among the leaked tetranucleate products of meiosis, however, the distribution pattern of *cenV-GFP* was consistent with proper segregation at meiosis I but random segregation at meiosis II (Fig. 5c). We also found that tagging chromosomal ScSGO1 with a 13-Myc tag at its carboxy terminus, which by itself yields no detectable defects in mitotic growth or meiosis I, resulted in impaired segregation at meiosis II (34% non-disjunction indicating 68% random segregation) (Fig. 5d). Moreover, the Myc-tagged ScSgo1 cells showed frequent separation of sister centromeres at late meiotic anaphase I (Fig. 5e), indicating that centromeric cohesion was not properly protected. Together, these results support the idea that ScSgo1 has a crucial role in protecting centromeric cohesion throughout meiosis I, thereby

ensuring normal progression to meiosis II, as does fission yeast Sgo1.

Conservation of shugoshin among eukaryotes

Our BLAST searches identified only three Sgo1-like proteins, all in fungi: *S. pombe* Sgo2, *S. cerevisiae* ScSgo1 and *N. crassa* B23G1.060. As we found two conserved regions among these proteins, we used the Block Maker and MAST programs^{29,30} to search for related proteins under conditions of two-block sequences. This approach yielded several candidate proteins from various eukaryotes including fly, nematode, plant, mouse and human (Fig. 6). Notably, the list included *Drosophila* MEI-S332, a previously characterized protein essential for preserving centromeric cohesion in meiosis¹⁷, although the similarity score is marginal (*E*-value = 10). All other proteins in the list show a short stretch of similarity in the C-terminal basic regions, whereas the primary sequences in the first block are not conserved except that they all contain a putative coiled-coil. The space and sequences between these two blocks diverge among the proteins. As these blocks were previously identified to be important for MEI-S332 function³¹, we investigated the significance of the conserved regions in Sgo1. We changed several amino acids individually within these similarity blocks to alanines and investigated the function of the mutant proteins *in vivo* (Supplementary Fig. 6). We found that three conserved amino acids known to be critical for MEI-S332 function³¹ were also required for Sgo1 function (N13, V34 and S384 in MEI-S332; N29, I50 and S294 in Sgo1) (Fig. 6, marked as arrowheads). Other conserved amino acids within the second block (P293, R296, K298, L299, R300 in Sgo1) were again all required for Sgo1 function (Fig. 6, asterisks), whereas non-conserved residue T297 could be changed to alanine without loss of function (Fig. 6, circle). These results underscore that the marginal

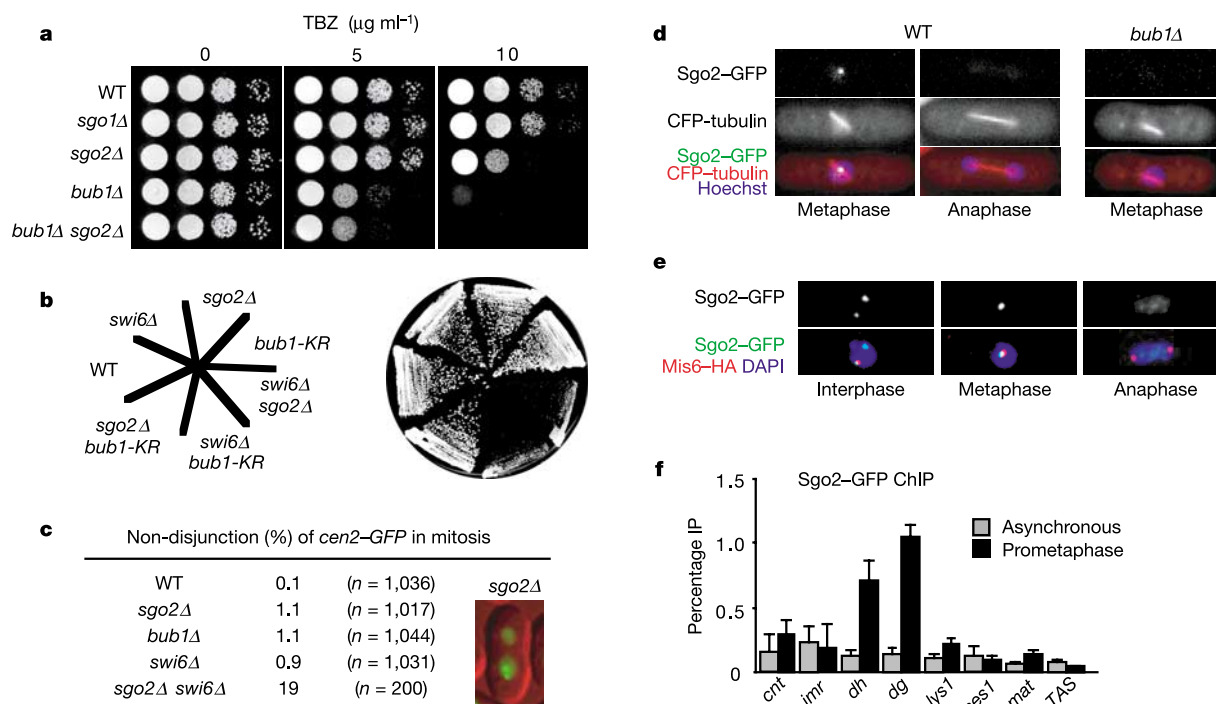


Figure 4 Sgo2 has a crucial role in mitotic division at the kinetochore. **a**, Serial dilutions of the indicated cultures were spotted onto YEA plates containing 0, 5 or 10 μg ml⁻¹ TBZ and incubated at 30 °C for 3 days. **b**, The indicated strains were streaked on YEA plates and incubated at 30 °C for 3 days. **c**, The frequency of non-disjunction of *cen2-GFP* was monitored in the indicated proliferating cells. **d**, Sgo2-GFP (green) was detected at metaphase in wild-type and *bub1Δ* cells expressing CFP-Atb2 for the visualization of

spindles (red). DNA was stained with Hoechst (blue). A wild-type cell at anaphase is also shown. **e**, The *sgo2⁺-GFP mis6⁺-HA* cells were fixed and stained with anti-GFP and anti-HA antibodies. Note that the Sgo2-GFP signal near Mis6-HA is faint or undetectable in interphase cells but is obvious in metaphase cells. **f**, A ChIP assay was used to measure Sgo2-GFP levels throughout the indicated chromosome sites in cells arrested at prometaphase or in asynchronous cells.

structural similarity found between *S. pombe* Sgo1 and the other proteins in various eukaryotes is significant. Plants and mammals carry two shugoshin-like proteins, suggesting that the function of shugoshin may have diverged to accomplish mitosis and meiosis, as in fission yeast.

Discussion

Our results provide some mechanistic insight into how shugoshin protects centromeric cohesion in meiosis. Previously we have suggested that Rec8 complexes enriched at the pericentromeric regions are crucial for preserving centromeric cohesion through meiosis I (ref. 18). Here we show that the protector protein Sgo1 also localizes at pericentromeric regions during meiosis I (Fig. 3d), consistent with the notion that Sgo1 protects precisely those Rec8 complexes that preserve centromeric cohesion. We found that Rec8 localization is not dependent on Sgo1 and vice versa (Fig. 2e, not shown). This independency of localization ensures that the mechanism protects Rec8 only at centromeres and not along chromosome arm regions. We suggest that shugoshin shields Rec8 physically from separase action or counteracts it. In this regard, we found that strong overexpression of Sgo1 moderately interferes with mitotic growth even in the absence of Rec8 expression (not shown), and that mild expression of Sgo1 kills a *cut1* mutant³² even at the permissive

temperature for the *cut1* allele (Supplementary Fig. 7). These results suggest that Sgo1 itself might have some ability to counteract separase function *in vivo*, although further analysis is necessary to address the precise mechanism by which shugoshin counteracts separase at the centromeres. We have shown that the co-expression of Sgo1 and Rec8 leads to the inability of sister chromatids to separate in mitosis (Fig. 1), suggesting that Sgo1 itself has the ability to protect Rec8 from degradation. However, in budding yeast cells, which express only a single shugoshin homologue, expression of Rec8 does not lead to a block in mitosis unless Spo13 is also expressed^{14,15}. This observation suggests that Spo13 is a potential meiotic activator of budding yeast shugoshin. As *S. pombe* cells expressing Rec8 instead of Rad21 exhibit normal mitotic growth (Fig. 1a), Sgo2 has no obvious activity in protecting Rec8 during mitosis. Thus, *S. pombe* Sgo1 appears to be well developed as a specialized Rec8 protector, obviating the need for meiosis-specific activators of shugoshin in fission yeast, and potentially explaining why Spo13 is not conserved.

Although fission yeast Sgo2 has some minor involvement in meiosis, it appears to function primarily in mitosis, whereas Sgo1 is dispensable for mitosis. Previous studies in *Drosophila* revealed that MEI-S332, presumably the only shugoshin in this organism, localizes to centromeres in mitosis and may have some role in strengthening cohesion³³. Here we find that the single budding yeast shugoshin has an important role in mitotic chromosome segregation, as the shugoshin mutant shows obvious chromosome instability during proliferation (Fig. 5b). Fission yeast *sgo2Δ* cells also show a modest defect in mitotic chromosome segregation and it becomes marked if an HP1 homologue, Swi6, is simultaneously depleted (Fig. 4b, c). Therefore, Sgo2 has an important role at mitotic kinetochores in a redundant capacity with centromeric heterochromatin. As heterochromatin is required to recruit large amounts of cohesin to centromeres^{34,35}, our results suggest that mitotic shugoshin may also have a close functional relationship with cohesin. One proposal would be that Bub1 senses a lack of

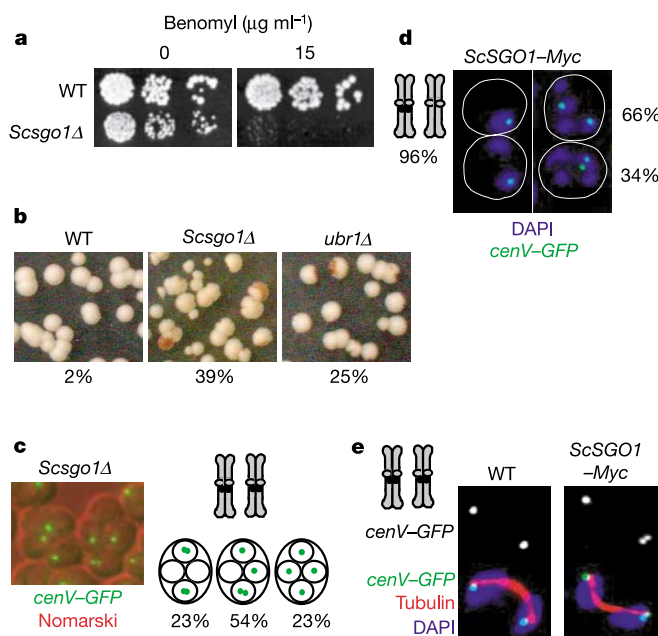


Figure 5 Analysis of budding yeast shugoshin ScSgo1. **a**, Serial dilutions of the indicated cultures were spotted onto YPD plates containing 0 or 15 $\mu\text{g ml}^{-1}$ benomyl. **b**, Chromosome loss was analysed in wild-type (WT) and *Scsgo1Δ* mutants by a colony-sectoring assay. The *ubr1Δ* mutant was used as a positive control⁴⁰. The frequency of sectoring colonies is shown at the bottom ($n > 120$). **c**, Examples of segregation of *cenV-GFP* in *Scsgo1Δ* tetrads. The segregation patterns in tetrads were classified mostly as one of the three shown at the right ($n = 200$). **d**, *ScSGO1-Myc* diploids were induced to synchronous meiosis and examined for the segregation of *cenV-GFP* marked on one of two homologues at meiosis I and meiosis II. The cells mostly underwent a reductional segregation pattern at meiosis I (96%, $n = 207$), whereas there was a high incidence of non-disjunction at meiosis II (34%, $n = 322$). **e**, Cells marked with *cenV-GFP* on both homologues were induced to meiosis and counter-stained with anti-tubulin antibody and DAPI. Cells at late anaphase I were examined for dots of *cenV-GFP* expression. *ScSGO1-Myc* cells frequently showed split *cenV-GFP* expression dots at either pair of sister chromatids (72%, $n = 138$), whereas control wild-type cells did not (<2%, $n = 106$).

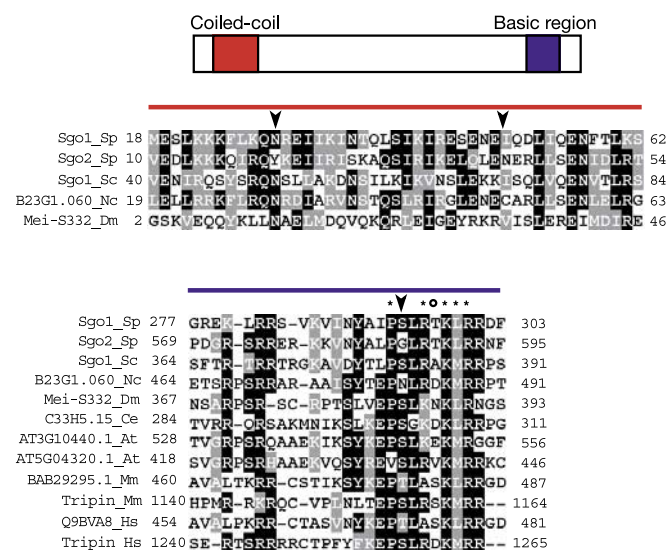


Figure 6 Alignment of the amino-terminal coiled-coil regions and C-terminal basic regions of shugoshin-like proteins in various organisms. The primary sequences of the N-terminal regions of Sgo1 are conserved among *S. pombe* (Sp; Sgo1 and Sgo2), *S. cerevisiae* (Sc; ScSgo1) and *N. crassa* (Nc; B23G1.060), whereas the sequences in other species, including MEI-S332, are not conserved, although all carry the putative coiled-coil motif (predicted by COILS program⁴⁴). Dm, *Drosophila melanogaster*; Ce, *Caenorhabditis elegans*; At, *Arabidopsis thaliana*; Mm, *Mus musculus*; Hs, *Homo sapiens*. See text for definition of arrowheads, asterisks and circle.

kinetochore–microtubule attachments or tension and thus stabilizes Sgo2 to maintain cohesion at centromeres. However, we could not detect any decrease of the cohesin Rad21 from centromeres in *sgo2Δ* cells even when the spindles were destabilized by the *nda3* mutation so that all cells were arrested at prometaphase by spindle checkpoint (T.S.K. and Y.W., unpublished observations). The detailed analysis of shugoshin function at mitotic kinetochores will be an intriguing subject of future study.

Our studies revealed that the conserved kinetochore kinase Bub1 is required for the centromeric localization of both mitotic and meiotic shugoshins, suggesting that Bub1 might regulate loading and/or maintenance of shugoshin at kinetochores. Although it is unclear whether shugoshin is a direct substrate of Bub1, the coincident localization of these proteins at metaphase kinetochores during mitosis and meiosis I suggests a possible close interaction between these proteins. Notably, *sgo1Δ* and *sgo2Δ* mutants share many if not all of the defects of *bub1Δ* cells in meiosis and mitosis, respectively (Figs 2d and 4b, c). Thus, we propose that shugoshin is a crucial target of Bub1 function at kinetochores. As mutations in human homologues of Bub1 have been found in subtypes of colorectal cancer that exhibit chromosome instability³⁶, our results suggest that human shugoshin may be a potential oncoprotein. In this regard, it is intriguing that the human shugoshin-like protein Q9BVA8 was recently identified as an antigen whose level is elevated in most breast cancers³⁷. Chromosome segregation in meiosis is also clinically important, as failures in this process result in aneuploidy, a major cause of miscarriage and birth defects in humans³⁸. In conclusion, we have identified a novel protein family, shugoshin, that protects sister chromatid cohesion proteins at centromeres during meiosis. Moreover, we have also established that shugoshin has a crucial role at kinetochores in guarding against chromosome instability during mitosis. This work provides a new model for considering the evolution of meiosis and an integrated understanding of eukaryotic chromosome segregation. □

Methods

Screening of the Rec8 protector

We searched for a gene that is toxic only when co-expressed with Rec8 in vegetative cells. The Rec8 coding sequence fused with GFP was cloned under the thiamine-repressible *nmt1* promoter into pREP82 (*ura4⁺* marker), to construct pREP82–*rec8⁺*–GFP. We used an *S. pombe* complementary DNA library, which was constructed using messenger RNA prepared from meiotic cells, and pREP3 vector (*nmt1* promoter, *LEU2⁺* marker) (Y. Akiyoshi and Y.W., unpublished observations). The *leu1 ura4-D18* cells carrying pREP82–*rec8⁺*–GFP were transformed with the cDNA library, spread on agar plates containing thiamine (promoter off) and incubated at 30 °C for 3 days. The colonies were then replicated onto two thiamine-free plates: one containing 5'-FOA (5'-fluoro-orthoic acid) and uracil where only cells that drop the plasmid pREP82–*rec8⁺*–GFP can grow (thereby expressing a library clone alone), and the other without 5'-FOA and uracil (allowing co-expression of *rec8⁺*–GFP and a library clone). We added phloxine B—a drug that stains dead cells red—to both plates, thereby highlighting dying colonies. After incubation for two days, the colonies exhibiting only dead cells on the co-expression plate were picked up, and the library-derived plasmids were recovered and analysed.

Schizosaccharomyces pombe strains

All strains used are listed in Supplementary Table 1. Deletion and GFP- or Flag-tagging of endogenous *sgo1⁺* and *sgo2⁺* were performed by a polymerase chain reaction (PCR)-based gene-targeting method³⁹. *sgo1⁺*–Flag–GFP was created by inserting GFP at the C terminus of the PCR-amplified *sgo1⁺*–Flag, and integrated at the endogenous *sgo1* locus. We further replaced the endogenous promoter of *sgo1⁺* with the *nmt1* promoter to generate *Pnmt*–*sgo1⁺* or *Pnmt*–*sgo1⁺*–Flag–GFP by the PCR-based gene targeting method³⁹. We abbreviate the tagged protein to Sgo1–GFP or Sgo1–Flag, depending on the purpose. We used a *mei4Δ* mutation to arrest meiotic cells before meiosis I (close to late prophase in meiosis I) and a *mes1Δ* mutation to arrest cells after meiosis I, as described⁸.

Chromatin immunoprecipitation assays

We used diploid *sgo1⁺*–Flag–GFP cells for ChIP assays with Sgo1. To achieve a highly synchronous culture, we replaced the endogenous *slp1⁺* (*CDC20* homologue) promoter with the *rad21⁺* promoter, which is not active during meiosis, to arrest the cells at metaphase I. The cells were incubated in nitrogen-depleted medium for 12 h at 30 °C with the result that approximately 60% of the cells were arrested at metaphase I. For ChIP with Sgo2, *nda3-KM311 sgo2⁺*–GFP cells were grown at 30 °C, and then shifted to 18 °C. After incubation for 8 h most of the cells were arrested at prometaphase. The cells were fixed with 3% paraformaldehyde at 18 °C for 30 min and extracts were prepared. ChIP assays

were carried out as described previously³⁵. The sequences of primers used have been described previously³⁰ except for the *mat* primers 5'-GTATGTGGAACAAGAGAG-3' and 5'-CTCGCCTGCTTACATTTAAGG-3'.

Preparation of anti-Sgo1 antibodies

The *sgo1⁺* ORF was amplified by PCR from an *S. pombe* cDNA library, and inserted into plasmids pGEX4T-2 (Pharmacia Biotech) and pET-19b (Novagen) for the production of recombinant proteins glutathione S-transferase (GST)–Sgo1 and His-tagged Sgo1, respectively. GST–Sgo1 was used to immunize a rabbit, and the raised antibodies were purified by His-tagged Sgo1 as previously described¹³.

Immunostaining

To stain endogenous Sgo1, wild-type diploid cells cultured for 5 h in MM-N medium were fixed with 3% formaldehyde and stained by the method described previously¹³. Sgo1 was detected using rabbit anti-Sgo1 antibodies and Alexa-488-conjugated anti-rabbit antibody (Molecular Probes). Tubulin was detected using the mouse anti-tubulin antibody TAT-1 (a gift from K. Gull) and Cy3-tagged anti-mouse antibody (Chemicon). For detecting GFP-tagged proteins, we used mouse anti-GFP antibody (Roche) and BODIPY FL-conjugated anti-mouse antibody (Molecular Probes). Mis6–haemagglutinin (HA) was detected with rabbit anti-HA antibody Y-11 (Santa Cruz) and Alexa-488-conjugated anti-rabbit antibody. Cells were counterstained with DAPI (4,6-diamidino-2-phenylindole) to visualize DNA.

Co-immunoprecipitation

Cells of strain *Padh*–*rec8⁺*–3HA *Pnmt41*–*sgo1⁺*–FLAG–GFP and control *Padh*–*rec8⁺*–3HA cells were cultured without thiamine for 15 h at 30 °C, collected, and extracts were prepared. To liberate chromatin-bound proteins, we treated the extracts with DNase I. After clarifying the extracts by centrifugation, the Sgo1–Flag–GFP protein was immunoprecipitated with anti-Flag antibody M2 (Sigma). Rec8–3HA and Sgo1–Flag–GFP were detected by anti-HA antibody Y-11 and anti-Flag antibody M2, respectively.

Analysis of budding yeast

All strains used are listed in Supplementary Table 1. The chromosome loss assay was carried out as described previously⁴⁰. The *ScSGO1* gene was deleted or epitope-tagged using PCR-generated cassettes⁴¹. Correct gene targeting was checked by PCR. *URA3*–GFP dots marking chromosome V (*cenV*–GFP) were described previously⁶. Sporulation was induced by incubating cultures of diploid cells at 30 °C as described previously⁴². *In situ* immunofluorescence was performed as described⁴².

Received 24 October; accepted 19 December 2003; doi:10.1038/nature02312.

Published online 18 January 2004.

- Nasmyth, K. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu. Rev. Genet.* **35**, 673–745 (2001).
- Koshland, D. E. & Guacci, V. Sister chromatid cohesion: the beginning of a long and beautiful relationship. *Curr. Opin. Cell Biol.* **12**, 297–301 (2000).
- Uhlmann, F. Chromosome cohesion and separation: from men and molecules. *Curr. Biol.* **13**, R104–R114 (2003).
- Lee, J. Y. & Orr-Weaver, T. L. The molecular basis of sister-chromatid cohesion. *Annu. Rev. Cell Dev. Biol.* **17**, 753–777 (2001).
- Hirano, T. The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. *Genes Dev.* **16**, 399–414 (2002).
- Klein, F. et al. A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. *Cell* **98**, 91–103 (1999).
- Parisi, S. et al. Rec8p, a meiotic recombination and sister chromatid cohesion phosphoprotein of the Rad21p family, conserved from fission yeast to humans. *Mol. Cell. Biol.* **19**, 3515–3528 (1999).
- Watanabe, Y. & Nurse, P. Cohesin Rec8 is required for reductional chromosome segregation at meiosis. *Nature* **400**, 461–464 (1999).
- Pasierbek, P. et al. A *Caenorhabditis elegans* cohesion protein with functions in meiotic chromosome pairing and disjunction. *Genes Dev.* **15**, 1349–1360 (2001).
- Eijpe, M., Offenberg, H., Jessberger, R., Revenkova, E. & Heyting, C. Meiotic cohesin REC8 marks the axial elements of rat synaptonemal complexes before cohesins SMC1β and SMC3. *J. Cell Biol.* **160**, 657–670 (2003).
- Stoop-Myer, C. & Amon, A. Meiosis: Rec8 is the reason for cohesion. *Nature Cell Biol.* **1**, E125–E127 (1999).
- Buonomo, S. B. et al. Disjunction of homologous chromosomes in meiosis I depends on proteolytic cleavage of the meiotic cohesin Rec8 by separin. *Cell* **103**, 387–398 (2000).
- Kitajima, T. S., Miyazaki, Y., Yamamoto, M. & Watanabe, Y. Rec8 cleavage by separase is required for meiotic nuclear divisions in fission yeast. *EMBO J.* **22**, 5643–5653 (2003).
- Shonn, M. A., McCarroll, R. & Murray, A. W. Spo13 protects meiotic cohesin at centromeres in meiosis I. *Genes Dev.* **16**, 1659–1671 (2002).
- Lee, B. H., Amon, A. & Prinz, S. Spo13 regulates cohesin cleavage. *Genes Dev.* **16**, 1672–1681 (2002).
- Blower, M. D. & Karpen, G. H. The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nature Cell Biol.* **3**, 730–739 (2001).
- Kerrebrock, A. W., Moore, D. P., Wu, J. S. & Orr-Weaver, T. L. MEI-5332, a *Drosophila* protein required for sister-chromatid cohesion, can localize to meiotic centromere regions. *Cell* **83**, 247–256 (1995).
- Kitajima, T. S., Yokobayashi, S., Yamamoto, M. & Watanabe, Y. Distinct cohesin complexes organize meiotic chromosome domains. *Science* **300**, 1152–1155 (2003).
- Toth, A. et al. Functional genomics identifies monopolin: a kinetochore protein required for segregation of homologs during meiosis I. *Cell* **103**, 1155–1168 (2000).
- Yokobayashi, S., Yamamoto, M. & Watanabe, Y. Cohesins determine the attachment manner of kinetochores to spindle microtubules at meiosis I in fission yeast. *Mol. Cell. Biol.* **23**, 3965–3973 (2003).
- Yamamoto, A. & Hiraoka, Y. Monopolar spindle attachment of sister chromatids is ensured by two

- distinct mechanisms at the first meiotic division in fission yeast. *EMBO J.* **22**, 2284–2296 (2003).
22. Mata, J., Lyne, R., Burns, G. & Bähler, J. The transcriptional program of meiosis and sporulation in fission yeast. *Nature Genet.* **32**, 143–147 (2002).
23. Watanabe, Y., Yokobayashi, S., Yamamoto, M. & Nurse, P. Pre-meiotic S phase is linked to reductional chromosome segregation and recombination. *Nature* **409**, 359–363 (2001).
24. Saitoh, S., Takahashi, K. & Yanagida, M. Mis6, a fission yeast inner centromere protein, acts during G1/S and forms specialized chromatin required for equal segregation. *Cell* **90**, 131–143 (1997).
25. Ekwall, K. *et al.* The chromodomain protein Swi6: a key component at fission yeast centromeres. *Science* **269**, 1429–1431 (1995).
26. Bernard, P., Maure, J. F. & Javerzat, J. P. Fission yeast Bub1 is essential in setting up the meiotic pattern of chromosome segregation. *Nature Cell Biol.* **3**, 522–526 (2001).
27. Yamaguchi, S., Decottignies, A. & Nurse, P. Function of Cdc2p-dependent Bub1p phosphorylation and Bub1p kinase activity in the mitotic and meiotic spindle checkpoint. *EMBO J.* **22**, 1075–1087 (2003).
28. Bernard, P., Hardwick, K. & Javerzat, J. P. Fission yeast bub1 is a mitotic centromere protein essential for the spindle checkpoint and the preservation of correct ploidy through mitosis. *J. Cell Biol.* **143**, 1775–1787 (1998).
29. Henikoff, S., Pietrokovski, S. & Henikoff, J. G. Superior performance in protein homology detection with the Blocks Database servers. *Nucleic Acids Res.* **26**, 309–312 (1998).
30. Bailey, T. L. & Gribskov, M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**, 48–54 (1998).
31. Tang, T. T., Bickel, S. E., Young, L. M. & Orr-Weaver, T. L. Maintenance of sister-chromatid cohesion at the centromere by the *Drosophila* MEI-S322 protein. *Genes Dev.* **12**, 3843–3856 (1998).
32. Kumada, K. *et al.* Cut1 is loaded onto the spindle by binding to Cut2 and promotes anaphase spindle movement upon Cut2 proteolysis. *Curr. Biol.* **8**, 633–641 (1998).
33. LeBlanc, H. N., Tang, T. T., Wu, J. S. & Orr-Weaver, T. L. The mitotic centromeric protein MEI-S332 and its role in sister-chromatid cohesion. *Chromosoma* **108**, 401–411 (1999).
34. Bernard, P. *et al.* Requirement of heterochromatin for cohesion at centromeres. *Science* **294**, 2539–2542 (2001).
35. Nonaka, N. *et al.* Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. *Nature Cell Biol.* **4**, 89–93 (2002).
36. Cahill, D. P. *et al.* Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300–303 (1998).
37. Scanlan, M. J. *et al.* Humoral immunity to human breast cancer: antigen definition and quantitative analysis of mRNA expression. *Cancer Immun.* **1**, 4 (2001).
38. Hassold, T. & Hunt, P. To err (meiotically) is human: the genesis of human aneuploidy. *Nature Rev. Genet.* **2**, 280–291 (2001).
39. Bähler, J. *et al.* Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14**, 943–951 (1998).
40. Rao, H., Uhlmann, F., Nasmyth, K. & Varshavsky, A. Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. *Nature* **410**, 955–959 (2001).
41. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).
42. Rabitsch, K. P. *et al.* Kinetochores recruitment of two nucleolar proteins is required for homolog segregation in meiosis I. *Dev. Cell* **4**, 535–548 (2003).
43. Glynn, J. M., Lustig, R. J., Berlin, A. & Chang, F. Role of bud6p and tealp in the interaction between actin and microtubules for the establishment of cell polarity in fission yeast. *Curr. Biol.* **11**, 836–845 (2001).
44. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. P. Cooper for critical reading of the manuscript; S. Hauf, M. Ohsugi and R. Watanabe for suggestions; J. P. Javerzat, F. Chang, T. Toda, M. Yanagida and P. Nurse for strains and plasmids of fission yeast; and F. Klein, K. P. Rabitsch, K. Nasmyth, M. Longtine, A. Shinohara and T. Maeda for strains and methods of budding yeast. We appreciate the support of M. Yamamoto and all members of his laboratory for their help. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.W. (ywatanab@ims.u-tokyo.ac.jp).

The formation of Kuiper-belt binaries through exchange reactions

Yoko Funato¹, Junichiro Makino², Piet Hut³, Eiichiro Kokubo⁴
& Daisuke Kinoshita⁵

¹General Systems Studies, University of Tokyo, Komaba, Meguro-ku, Tokyo 153, Japan

²Department of Astronomy, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

³Institute for Advanced Study, Princeton, New Jersey 08540, USA

⁴National Astronomical Observatory, Osawa, Mitaka-shi, Tokyo 180, Japan

⁵National Central University, 300 Jungda Road, Chung-Li 32054, Taiwan

Recent observations^{1–8} have revealed that an unexpectedly high fraction—a few per cent—of the trans-Neptunian objects (TNOs) that inhabit the Kuiper belt are binaries. The components have roughly equal masses, with very eccentric orbits that are wider than a hundred times the radius of the primary. Standard theories of binary asteroid formation tend to produce close binaries with circular orbits, so two models have been proposed^{9,10} to explain the unique characteristics of the TNOs. Both models, however, require extreme assumptions regarding the size distribution of the TNOs. Here we report a mechanism that is capable of producing binary TNOs with the observed properties during the early stages of their formation and growth. The only required assumption is that the TNOs were initially formed through gravitational instabilities¹¹ in the protoplanetary dust disk. The basis of the mechanism is an exchange reaction in which a binary whose primary component is much more massive than the secondary interacts with a third body, whose mass is comparable to that of the primary. The low-mass secondary component is ejected and replaced by the third body in a wide but eccentric orbit.

The TNO binary with the best-known orbital elements, 1998WW31, has a mass ratio $m_2/m_1 \approx 0.7$, eccentricity $e \approx 0.8$, and semi-major axis $a \approx 2 \times 10^4 \text{ km}^2$. The inferred radii of the components are $r_1 \approx 1.1r_2 \approx 10^2 \text{ km}$, hence $a/r_1 > 10^2$, in stark contrast to main belt asteroid binaries where $m_2/m_1 \ll 1$, $e \approx 0$, and $a/r_1 \lesssim 10$. Asteroid binaries in the main belt are probably formed by collisions¹², as in the leading theory for the formation of the Moon^{13,14}, which naturally explains the extreme mass ratios and tight near-circular orbits^{15,16}.

To find a completely different process for TNO binaries, we can borrow ideas from dynamical binary formation in star clusters, where many roughly equal-mass binaries are found. There, we know of three formation paths: (1) two-body dissipational encounters¹⁷; (2) three-body binary formation¹⁸; and (3) exchange reactions¹⁸. Figure 1 schematically shows these paths.

In the case of TNOs and asteroids, path (1) corresponds to the standard theoretical conditions for binary formation: tidal disruption and giant impact during a close encounter (Fig. 1a, b). Such conditions indeed occur: each TNO has grown through accretion, and much of accretion has happened through collisions with an object comparable in mass to that of the growing TNO itself^{19,20}.

Path (2) (Fig. 1c) would require a near-simultaneous encounter of three massive objects with low enough velocities to allow an appreciable chance of leaving two of the objects bound. For this to work, the random velocities of the most massive objects should be significantly lower than their Hill velocities, $v_H \equiv R_H \Omega(R)$, where R_H is the Hill radius of the asteroid at a distance R from the Sun and $\Omega(R)$ is its orbital angular velocity. This path could work if there are about 10^5 100-km-sized objects embedded in a sea of small ($< 1 \text{ km}$) objects¹⁰. This assumption, however, is at odds with Goldreich and Ward's theory for the formation of planetesimals¹¹ through gravita-

tional instability, and it is hard to see how objects in the Kuiper belt could form from non-gravitational coagulation, because the time-scales are far too long²¹. In contrast, the gravitational instability theory predicts the size of the initial bodies to be 10–100 km. Starting with these bodies would make path (2) ineffective, because the velocity dispersion would be higher than the Hill velocity^{19,22}.

A combination of paths (1) and (2) has been proposed⁹. Weiden-schilling studied how a third massive body could capture a collision product of two massive bodies if the third body were near enough during the time of the collision. This mechanism seems unlikely to work, because it requires a number density of massive objects that is about two orders of magnitude higher than the value consistent with observations¹⁰.

Another mechanism, based on the dynamical friction from smaller bodies, in which a hyperbolic encounter between two massive bodies produces a bound orbit, has been proposed¹⁰. As mentioned above, the gravitational instability theory for the formation of planetesimals¹¹ would exclude the existence of such a sea of small objects. Path (3) (Fig. 1d) can operate on the binaries formed through path (1) (Fig. 1b), so we should check whether paths (1) and (3) together produce the correct binaries in the correct numbers.

Consider a relatively massive TNO primary with mass m_1 , in a binary with a secondary with mass $m_2 \ll m_1$. If the binary encounters another body with mass $m_3 \approx m_1$, the most likely result is an exchange reaction, in which the incoming object replaces the original secondary²³. Figure 2 shows an example of such a reaction. If the incoming velocity of the third body is small, the binding energy of the resulting binary is comparable to or larger than that of the initial binary, because otherwise the secondary could not escape. In practice, the change in binding energy is relatively small, as will be shown later. Hence $m_1 m_2 / a_0 \approx m_1^2 / a$, where a is the new semi-major axis after the exchange. This implies $a/a_0 \approx m_1/m_2 \gg 1$.

For the exchange reaction to occur, the third body should approach the binary to a distance comparable to a_0 . The angular momentum carried away by the light body is small. Therefore, the periastron distance of the resulting binary is comparable to that of the initial parabolic orbit, in other words, to a_0 . Thus, the eccentricity is $e \approx 1 - a_0/a \approx 1 - m_2/m_1$.

We have run a series of numerical scattering experiments to

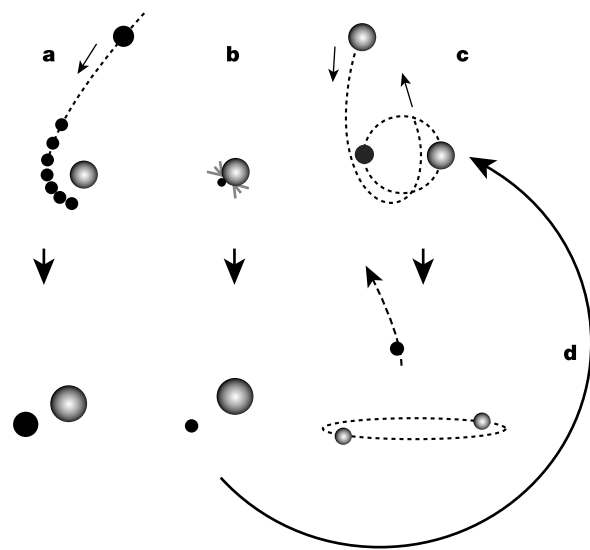


Figure 1 Paths of formation of binaries. Path **a** is the formation through tidal disruption of one object followed by coagulation of fragments during a close encounter with the other. Path **b** is a giant impact, where collision debris coagulates into a "moon". Path **c** is an exchange reaction, where the incoming body replaces one component of a binary. Path **d** is a combination of Paths **b** and **c**.

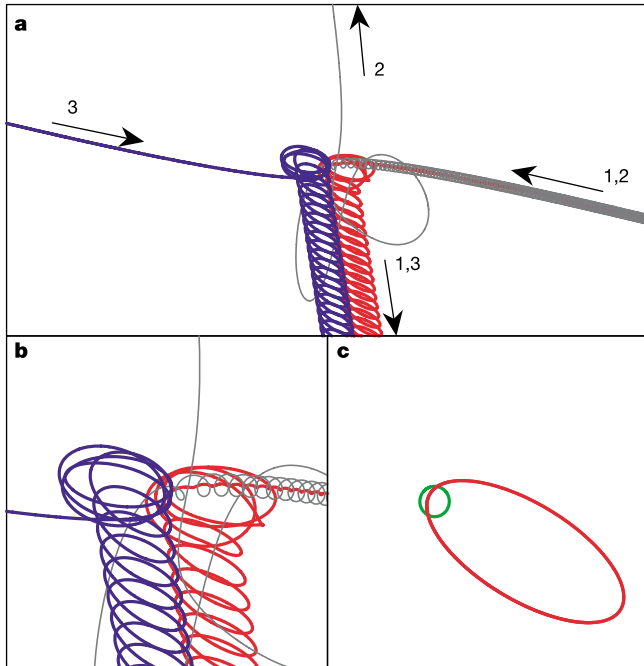


Figure 2 An example of a binary–single-body exchange interaction. Bodies 1 and 2 having masses $m_1 = 1$ and $m_2 = 0.1$, respectively, form a binary with an initially circular orbit. Body 3, with mass $m_3 = 1$, encounters the binary on an initially parabolic orbit (blue). In **a**, the whole scattering process is shown. **b**, The complex central interaction shown in more detail. **c**, The orbits of the initial (green) and final (red) binary, respectively. The final binary orbit is highly eccentric and much wider than the initial circular binary orbit.

obtain cross-sections for different outcomes of interactions between a binary and a single body (see Methods for details). The results show that 80% of the cross-sections corresponding to the change of the composition of the binary is accounted for by processes leading to the formation of binaries with two massive components.

In Fig. 3 the distribution for the semi-major axis is strongly peaked around $a \approx 20$. This result is in good agreement with the simple argument presented above. This peak arises from the exchanges which take place at the first binary–single-body encounter. The broad peak around $a \approx 10$ arises from ‘resonant encounters’, in which encounters of two objects occur more than twice²⁴. The eccentricity peaks at 0.95, as expected. In the case of 1998WW31 with $r_1 = 75$ km, our length unit corresponds to 1,500 km. In Figs 2 and 3 we attach this physical scale to give a concrete idea of the size. Figure 4 clearly shows that the orbital elements of 1998WW31 are consistent with the binary having formed through the processes modelled here.

Our second task is to check whether the exchange path is efficient enough to produce the observed binaries. Starting with TNOs formed through gravitational instability of the protoplanetary dust layer¹¹, the heaviest TNOs will accrete mass primarily through collisions with TNOs of comparable mass^{19,20}. Many of these collisions are of the ‘giant impact’ type, resulting in tight binaries containing components of very unequal mass (Fig. 1b). Let us estimate how many such binaries are formed and what fraction of them are converted to wide binaries through exchange.

We assume that one in three collisions between comparable TNOs gives rise to a binary. The numerical simulations of planetesimal collisions yield a sizeable formation rate of satellites of 43.3% (ref. 16). Because the relative velocity between objects in the TNO region is significantly lower than that assumed in their simulations, the corresponding binary formation rate in the TNO region is even higher. Conservatively, however, we have assumed a

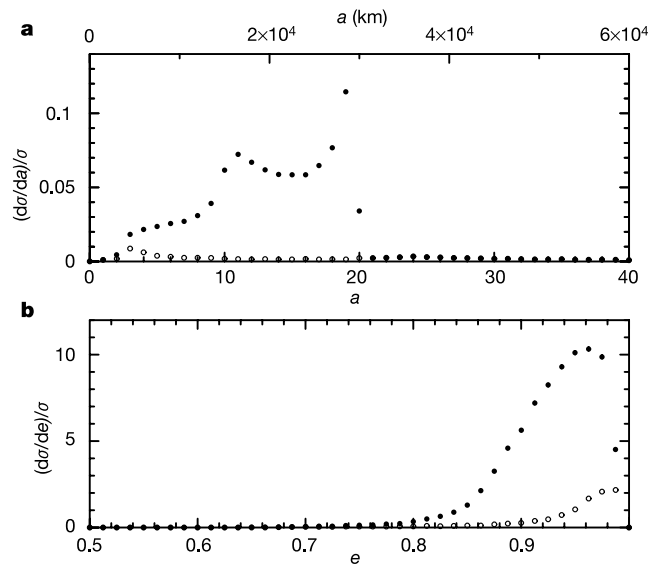


Figure 3 Normalized differential cross-sections for the formation of a ‘massive–massive’ binary. Normalized differential cross-sections for the event types (1), (3) and (5) in Methods are plotted with respect to the semi-major axis **a** (**a**), and eccentricity **e** (**b**) of the final binary. The physical units are given for reference at the top of the figure. The filled points are the total values, while the open circles are the contributions from the merger events (types (3) and (5) in Methods).

ratio of only 1/3.

In 2/3 of the cases, when no binary is produced, we have to wait for a typical time T until another collision occurs. When a binary is formed, the cross-section for subsequent interactions with a third body, taking into account gravitational focusing, is in proportion to a_0 . Therefore, our newly formed binary will undergo an exchange reaction on a timescale $(r/a_0)T \ll T$, leading to a significant increase in a . Strong three-body interactions will subsequently occur on a shorter timescale $(r/a)T \ll T$. As a result, the semi-major axis will shrink systematically, with a ‘thermal’ distribution $f(e) = 2e$ favouring high eccentricity¹⁸. When the orbit becomes small enough ($r/a \approx 0.03$), the chance for collisions in resonant encounters becomes significant²⁵. Let us assume that an exchange reaction turns a ‘giant impact’ binary into a binary with $a \approx 300r$. Each subsequent strong encounter will on average shrink a by a factor²⁶ of about 1.2. After a dozen encounters, $a \approx 30r$ and a collision is likely to occur. The timescale for each encounter to occur is $\sim (r/a)T$, and the waiting time for the last encounter in this series is $(1/30)T$. Summing this series, we get a total waiting time of $(T/30)/(1 - (1/1.2)) = 0.2T$ before a collision. If three bodies collide, or if the third body escapes, the resulting system may be a single body (with an assumed chance of 2/3) or a binary of very-unequal mass components (a chance of 1/3). If the merger product and the third body are still bound, we have an equal-mass wide binary. Under these assumptions, in 1/3 of the cases, the result is an equal-mass TNO binary with the observed properties for a period $\sim 0.2T$, compared to a 2/3 chance of ending up with a single TNO for a period of $\sim T$. Denoting by N_S and N_B the number of single bodies and the number of binaries, respectively, we have:

$$\frac{dN_B}{dt} = \frac{1}{3}N_S - \frac{1}{0.23}N_B$$

$$\frac{dN_S}{dt} = -\frac{dN_B}{dt}$$

if we measure time in units of T . For the stationary state we have $dN_B/dt = dN_S/dt = 0$, and $N_B = 0.1N_S$. Therefore, the predicted binary fraction is $\sim 10\%$. When accretion in the Kuiper-belt region

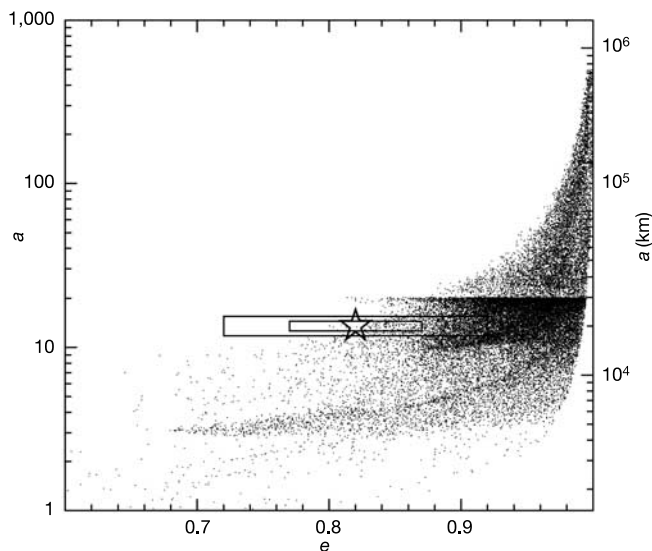


Figure 4 Distribution of orbital properties of 'massive-massive' binaries formed in our scattering experiments. Here a and e have the same meanings as in Fig. 3. Contributions from exchange reactions, event type (1) in Methods, are limited by energy conservation to $a \lesssim 20$, and give rise to the horizontal rim in the middle of the figure. Contributions involving mergers, event types (3) and (5), can lead to a values as high as the Hill radius $a \approx 10^3$. The star symbol shows the observed orbit for 1998WW31. Boxes around the star indicate the observational 1σ and 2σ error bars.

diminished, the number of single and binary objects was frozen, with a ratio similar to this steady-state value.

Although our arguments are approximate, it is clear that after cessation of the accretion stage, at least several per cent of the TNOs were accidentally left in such a binary phase. That more than 1% of the known TNOs are found to be in wide, roughly equal-mass binaries is thus a natural consequence of any accretion model, independent of the assumed parameters for the density and velocity dispersion of the protoplanetary disk or the duration of the accretion phase. We predict that future discoveries of TNO binaries will similarly show rough equal masses, large separations and high eccentricities. □

Methods

Numerical scattering experiments

We performed 775,541 runs of numerical integrations of binary-single-body encounters. We chose the relative velocity v_∞ between the intruder (single body) and the binary at infinity to be much smaller than the orbital velocity of the binary components. In this regime, the scattering cross-section σ is expressed as $\sigma \propto \sigma_0/v_\infty^2$, because of gravitational focusing²⁶. The parabolic orbit is the limiting case of $v_\infty \rightarrow 0$ where the nominal cross-section σ diverges, but σ_0 converges to a finite value.

We fixed the initial orbital parameters of the binary, and changed the relative orbit of the intruder to obtain scattering cross-sections. For encounters with finite v_∞ , one should randomly change the direction of v_∞ and also randomly select impact parameter vectors uniformly within a circle of given radius. This radius should be large enough so that encounters with impact parameter larger than this radius would not result in exchange or collision. For parabolic encounters, this procedure choice corresponds to choosing the periastron distance r_p from a uniform distribution with an upper limit and selecting random orientations for the orbital plane and direction of the periastron. As an upper limit, we chose $20a_0$, where a_0 is the semi-major axis of the initial binary. Initial separations between the intruder and the binary were $100a_0$.

We use units in which $G = m_1 = m_3 = a_0 = 1$, where G is the gravitational constant, and m_1 and m_3 are the masses of the heavier body of the initial binary and the intruder. The mass of the secondary component is $m_2 = 0.05$. The radii of the bodies are $r_1 = r_3 = 0.05$ and $r_2 = r_1(m_2/m_1)^{1/3} \approx 0.01842$. The initial binary is circular, and separations are 5–40 times the radius of the primary. These values are typical for main-belt binary asteroids, with $m_2/m_1 < 0.1$.

We terminated the numerical integration either when one of the three bodies escaped beyond the Hill radius, or when two bodies physically collided. For the 'escape' scenario we chose a Hill radius of $r_H = 100a_0$, because a typical value for a TNO with radius r , $m = 10^{21}$ g and $R = 40$ AU is $a = 1,000r$, and the typical semi-major axis of a close, circular planetesimal binary is $a_0 = 10r$. For the 'collision' scenario, the collision product was assumed to inherit the mass and motion of the centre of mass of the two bodies. We

adapted the radius to conserve the density. In both scenarios, we calculated the orbit of the remaining two bodies to determine the final state.

We integrated the equations of motion of the three bodies in cartesian coordinates. Integration was done with an 8th-order time-symmetric variable-stepsize hermite method. To cross-check the results, we also performed two other sets of scattering experiments, one using the scattering package of STARLAB²⁷, and the other using a 4th-order hermite scheme. In all cases, we used completely different programs for generating initial conditions, as well as for orbit integration and for the reduction of the results. All three experiments gave the same outcome, within the expected fluctuations associated with the finite number of runs.

There are six possible configuration changes that can take place as a result of a scattering event: (1) an exchange reaction resulting in a massive-massive binary; (2) an exchange reaction resulting in a massive-light binary; (3) a merger of bodies 1 and 2 resulting in a massive-massive binary; (4) a merger of bodies 1 and 3 resulting in a binary with large mass ratio (2.1:0.05); (5) a merger of bodies 2 and 3 resulting in a massive-massive binary; (6) no binary is left, after three-body merging or two-body merging followed by escape. The cross-section σ_0 for each event is as follows: (1) 12.1, (2) 1.3, (3) 0.9, (4) 1.3, (5) 0.9 and (6) 1.2, in our units. For our scattering experiments with $r_p < 20a_0$, we found a cross-section for preservation of the original binary of 1243.9 in our units, which implies that in 99% of our experiments no configuration change took place. We also found that for $r_p > 20a_0$ no merger or exchange occurred, so our cross-section measurements are complete.

Another logical possibility would be: (7) all three objects become unbound. However, for that to occur the total energy would have to be positive. In our case, the low relative velocity of the incoming third body guarantees that the total energy is negative, so (7) cannot occur.

The physical reason for the dominance of (1) is the statistical tendency for equipartition of kinetic energy, which implies much higher velocities for much lighter bodies and therefore a much larger chance of escape for those lighter bodies.

Received 28 March; accepted 29 December 2003; doi:10.1038/nature02323.

1. Burns, J. A. Two Bodies Are Better Than One. *Science* **297**, 942–943 (2002).
2. Veillet, C. *et al.* The binary Kuiper-belt object 1998 WW 31. *Nature* **416**, 711–713 (2002).
3. Margot, J.-L. Worlds of mutual motion. *Nature* **416**, 694–695 (2002).
4. Noll, K. S. *et al.* Detection of two binary trans-Neptunian objects, 1997 CQ₂₉ and 2000 CF₁₀₅, with the Hubble Space Telescope. *Astron. J.* **124**, 3424–3429 (2002).
5. Schaller, E. L. & Brown, M. E. A deep Keck search for binary Kuiper Belt objects. *DPS 35th Meeting Abstr.* no. 39.20 (American Astronomical Society, Washington DC, 2003).
6. Noll, K. S. *et al.* Discovery of transneptunian binaries with HST. *DPS 35th Meeting Abstr.* no. 49.70 (American Astronomical Society, Washington DC, 2003).
7. Noll, K. S. *et al.* 2001 QC₂₉₈. *IAU Circ. No.* 8034 (2002).
8. Noll, K. S. & Stephens, D. C. 1999 RZ₂₅₃. *IAU Circ. No.* 8143 (2003).
9. Weidenschilling, S. J. On the origin of binary transneptunian objects. *Icarus* **160**, 212–215 (2002).
10. Goldreich, P., Lithwick, Y. & Sari, R. Formation of Kuiper-belt binaries by dynamical friction and three-body encounters. *Nature* **420**, 643–646 (2002).
11. Goldreich, P. & Ward, W. R. The formation of planetesimals. *Astrophys. J.* **183**, 1051–1062 (1973).
12. Melrose, W. J. *et al.* in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolucci, P. & Binzel, R. P.) 273–287, (Lunar and Planetary Institute, Houston, 2003).
13. Hartmann, W. K. & Davis, D. R. Satellite-sized planetesimals and lunar origin. *Icarus* **24**, 504–515 (1975).
14. Kokubo, E., Ida, S. & Makino, J. Evolution of a circumterrestrial disk and formation of a single moon. *Icarus* **148**, 419–436 (2000).
15. Durda, D. D., Bottke, W. F., Asphaug, E. & Richardson, D. C. The formation of asteroid satellites: Numerical simulations using SPH and N-body models. *Bull. Am. Astron. Soc.* **33**, 1134 (2001).
16. Durda, D. D. *et al.* The formation of asteroid satellites in catastrophic impacts: Results from numerical simulations. *Lunar Planet. Sci. XXXIV Abstr.* no. 1943 (Lunar and Planetary Institute, Houston, 2003).
17. Fabian, A. C., Pringle, J. E. & Rees, M. J. Tidal capture formation of binary systems and X-ray sources in globular clusters. *Mon. Not. R. Astron. Soc.* **172**, 15–18 (1975).
18. Heggie, D. C. Binary evolution in stellar dynamics. *Mon. Not. R. Astron. Soc.* **173**, 729–787 (1975).
19. Kokubo, E. & Ida, S. Oligarchic growth of protoplanets. *Icarus* **131**, 171–178 (1998).
20. Makino, J., Fukushima, T., Funato, Y. & Kokubo, E. On the mass distribution of planetesimals in the early runaway stage. *New Astron.* **3**, 411–416 (1998).
21. Wetherill, G. W. Formation of the earth. *Annu. Rev. Earth Planet. Sci.* **18**, 205–256 (1990).
22. Kenyon, S. J. & Luu, J. X. Accretion in the Early Kuiper Belt. I. Coagulation and velocity evolution. *Astron. J.* **115**, 2125–2160 (1998).
23. Spitzer, L. *Dynamical evolution of globular clusters*. (Princeton Univ. Press, Princeton, NJ, 1987).
24. Hut, P. Hard binary-single star scattering cross sections for equal masses. *Astrophys. J. Suppl.* **55**, 301–317 (1984).
25. Hut, P. & Inagaki, S. Globular cluster evolution with finite-size stars—Cross sections and reaction rates. *Astrophys. J.* **298**, 502–520 (1985).
26. Heggie, D. C. & Hut, P. *The Gravitational Million-Body Problem* Ch. 23 (Cambridge Univ. Press, Cambridge, 2003).
27. McMillan, S. & Hut, P. Binary-single-star scattering. VI. Automatic determination of interaction cross sections. *Astrophys. J.* **467**, 348–358 (1996).

Acknowledgements We acknowledge comments on our manuscript by P. Goldreich, R. Rafikov, R. Sari, K. S. Noll and D. Durda.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.F. (funato@chianti.c.u-tokyo.ac.jp).

Statistical mechanics of a gas-fluidized particle

R. P. Ojha¹, P.-A. Lemieux¹, P. K. Dixon^{1*}, A. J. Liu² & D. J. Durian¹

¹Department of Physics and Astronomy, University of California, Los Angeles, California 90095-1547, USA and ²Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095-1569, USA

* Present address: Department of Physics, California State University, San Bernardino, California 92407, USA

Characterization of the microscopic fluctuations in systems that are far from equilibrium is crucial for understanding the macroscopic response. One approach is to use an 'effective temperature'—such a quantity has been invoked for chaotic fluids^{1,2}, spin glasses^{3,4}, glasses^{5,6} and colloids^{7,8}, as well as non-thermal systems such as flowing granular materials^{9–14} and foams¹⁵. We therefore ask to what extent the concept of effective temperature is valid. Here we investigate this question experimentally in a simple system consisting of a sphere placed on a fine screen in an upward flow of gas; the sphere rolls because of the turbulence it generates in the gas stream. In contrast to many-particle systems, in which it is difficult to measure and predict fluctuations, our system has no particle–particle interactions and its dynamics can be captured fully by video imaging. Surprisingly, we find that the sphere

behaves exactly like a harmonically bound brownian particle. The random driving force and frequency-dependent drag satisfy the fluctuation–dissipation relation, a cornerstone of statistical mechanics. The statistical mechanics of near-equilibrium systems is therefore unexpectedly useful for studying at least some classes of systems that are driven far from equilibrium.

To fluidize a granular system, energy must be injected, for example by mechanical agitation or an upflow of gas strong enough to counter gravity^{16,17}. The latter method is used industrially in gas-fluidized bed reactors¹⁸, where the medium is a catalyst powder and the gas is a reactant. The nature of the fluidization transition versus flow rate, and the mechanical properties of the fluidized medium, are topics that still pose a formidable research challenge; this is largely because the relative importance of collisional, cohesive, and gas-mediated interactions is not well understood. In part to isolate gas-mediated interactions, we constructed a gas-fluidized bed for large, easily visualized spheres (see Methods) to serve as a paradigm for more typical fluidized systems, where the collisions take place on length and timescales too small to be captured by video imaging.

The nature of fluctuations in other experimental driven systems has been explored previously^{19–24}. Gaussian speed distributions were found for rolling¹⁹ and sliding²⁰ particles, but not for all systems of vertically vibrated beads^{21–23}. In addition, Boltzmann energy distributions were found for a particle driven randomly by design²⁴. Here we focus on a single sphere rolling stochastically in an upflow of gas (see Supplementary video 1 and the Methods section). In contrast with previous work, we measure not only the speed and position statistics, but also the time-dependent dynamics. This unprecedented level of microscopic information allows direct contact with the fluctuation and drag forces, and enables a complete and detailed test of the applicability of near-equilibrium statistical mechanics to a far-from-equilibrium system.

The observed probability distributions for the ball speed and radial position are displayed in Fig. 1a and b. Evidently, the velocity and position components have the same gaussian form as a harmonically bound brownian particle in thermodynamic equilibrium. For thermal systems obeying the equipartition theorem, the mean-squared speed and radial position give the temperature and spring constant as $kT = m_e \langle v^2 \rangle / 2$ and $K = m_e \langle v^2 \rangle / \langle r^2 \rangle$, respectively. For our driven mechanical system, we may deduce an effective temperature and spring constant using these relations, but there is no guarantee that these quantities have any meaning. For instance, the same distributions in Fig. 1a and b could arise if the position varied sinusoidally versus time with a stochastic amplitude that changed only very slowly. But in fact, as we shall now demonstrate, the behaviour of the fluidized sphere is truly thermal in the usual statistical mechanical sense.

To verify the effective spring constant, we tilt the entire apparatus by angle θ away from the normal, so that the ball is subjected to

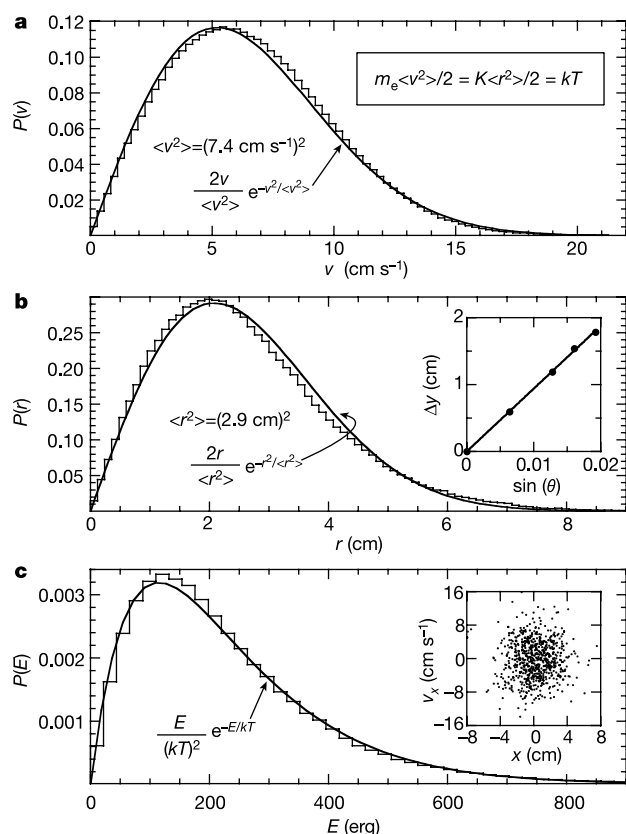


Figure 1 Speed (a), radial position (b), and energy (c) probability functions for a sphere that rolls stochastically owing to the turbulence it generates in an upflow of gas. The smooth curves are for a brownian particle in a two-dimensional harmonic trap. The inset of b shows the shift in average ball position versus the sine of the angle by which the entire apparatus is tilted away from horizontal; it verifies the spring constant extracted from the main plots. The inset of c shows a scatterplot of speed and position; it demonstrates absence of correlation.

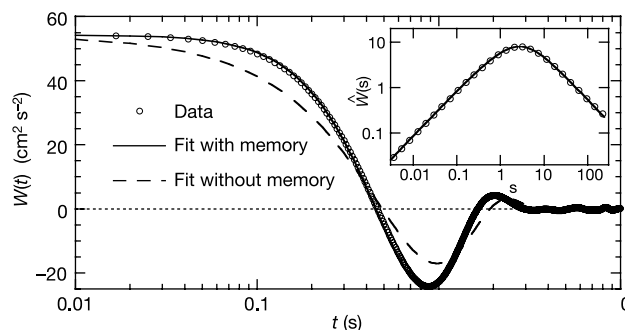


Figure 2 Velocity autocorrelation for the stochastically rolling sphere. The Laplace transform is shown in the inset. The solid (dashed) curve is a fit to the Langevin equation with (without) memory.

gravity, $mg\sin\theta$, within the plane. We find that the speed distribution is unaltered, and that the position distribution sags down the plane but remains gaussian. The shift in average position grows linearly with the sine of the tilt angle, as shown in the inset of Fig. 1b, and agrees well with expectation, $\Delta y = (mg\sin\theta)/K$, where $K = m_e\langle v^2 \rangle / \langle r^2 \rangle$. Thus, a true potential, here due to gravity, can simply be added to the 'harmonic potential' induced by gas flow, to describe the behaviour of the sphere. This confirms the existence of a harmonic confining potential, and the validity of equipartition for our driven system. As an aside, the root mean squared (r.m.s.) radial position of the ball is roughly one-fifth the sieve diameter, for any ball size, system size or air speed; therefore, we believe the harmonic potential is due to interaction of the ball's turbulent wake with the system boundary.

To verify the effective temperature, we compute the total mechanical energy, $E = m_e v^2/2 + Kr^2/2$, the sum of the kinetic and potential energies. The distribution shown in Fig. 1c agrees well with the expectation for a two-dimensional harmonically bound brownian particle, $P(E) = [E/(kT)^2] \exp(-E/kT)$, where $kT = m_e\langle v^2 \rangle/2$; this is the product of the Boltzmann factor and the density of states. The inset to Fig. 1c shows that phase space is sampled evenly, with no correlation between position and momentum. From Fig. 1c and its inset, we adduce that equal-energy states occur with equal probability.

Although Fig. 1 presents strong evidence of normal thermal behaviour, a definitive demonstration requires verification of the fluctuation–dissipation theorem²⁵. Specifically, we wish to determine whether the ball dynamics obey the Langevin equation:

$$m_e \mathbf{a}(t) = -K\mathbf{r}(t) + m_e \int_{-\infty}^t \Gamma(t-t') \mathbf{v}(t') dt' + \mathbf{F}_r(t) \quad (1)$$

where the random force, $\mathbf{F}_r(t)$, is related to the memory kernel for the drag, $\Gamma(t)$:

$$\langle \mathbf{F}_r(t) \cdot \mathbf{F}_r(t') \rangle = 2m_e kT \Gamma(t-t') \quad (2)$$

If equations (1) and (2) are satisfied, then the ball behaves as a perfectly thermal particle. To characterize the ball's stochastic motion, and to test equations (1) and (2), we compute the velocity autocorrelation function, $W(t) = \langle \mathbf{V}(0) \cdot \mathbf{V}(t) \rangle$, from position versus time data. The results in Fig. 2 indicate that the motion is slightly underdamped, with all correlations being lost within a few oscillations. The Laplace transform is shown in the inset. According to

equations (1) and (2), it should be

$$\hat{W}(s) = (kT/m_e) [s + \hat{\Gamma}(s) + \omega^2/s] \quad (3)$$

where $\hat{\Gamma}(s)$ is the Laplace transform of the memory kernel, $kT/m_e = \langle v^2 \rangle$, and $\omega^2 = K/m_e = \langle v^2 \rangle / \langle r^2 \rangle$. A simple form for the kernel, consistent with pressure fluctuations at higher Re (ref. 26), is an exponential: $\Gamma(t) = \Gamma_0 \gamma_0 \exp(-\gamma_0 t)$ and $\hat{\Gamma}(s) = \Gamma_0 \gamma_0 / (s + \gamma_0)$. Given $\langle v^2 \rangle$ and $\langle r^2 \rangle$ from Fig. 1, the best fit is shown in the inset, resulting in a drag amplitude and memory decay rate of $\Gamma_0 = 2.3 \text{ s}^{-1}$ and $\gamma_0 = 10 \text{ s}^{-1}$. The former is of order $(g/D)^{1/2}$, suggesting that rolling friction is responsible for the drag; the latter is comparable to the vortex shedding frequency, $0.15U/D$ (refs 26, 27), suggesting that the ball is driven randomly by the turbulence it generates in the gas (see discussion in Supplementary Information 2). The inverse transform of this fit, shown as the solid curve in Fig. 2, agrees very well with the velocity autocorrelation data. By contrast, if there is no memory ($\gamma_0 \rightarrow \infty$) then good agreement cannot be achieved.

Two more steps are required to complete the demonstration of the fluctuation–dissipation theorem. First, we must independently verify the drag force. Because the width of a resonance is related to dissipation, we sinusoidally tilt the entire apparatus at different frequencies. The average position of the ball, computed separately for each tilt angle, also varies sinusoidally. The amplitude, normalized by the sag for $\omega = 0$, and the phase of the response are shown in Fig. 3. The results are linear (identical for two different tilt amplitudes), and agree very well with the expectation based on the velocity autocorrelation fit with memory. Second, we must verify the random forcing. The spring and drag forces are now known, so we can simply subtract them from $m_e \mathbf{a}(t)$ to find $\mathbf{F}_r(t)$. As shown in the inset of Fig. 4, the distribution of random forces is gaussian, with a mean-squared value given by equation (2) as $\langle \mathbf{F}_r \cdot \mathbf{F}_r \rangle = 2m_e kT \Gamma_0 \gamma_0$. Moreover, the force autocorrelation function, shown in the main plot of Fig. 4, is very nearly exponential with an amplitude and decay rate given by equation (2). Thus the Langevin equation holds, with the drag force and the random force related by the fluctuation–dissipation relation. Altogether, the ball's behaviour is truly thermal, but with an effective temperature that is 10^{15} times larger than the ambient room temperature.

At first glance, it may not seem surprising that near-equilibrium statistical mechanics holds. The airflow speed is 50 times greater than the r.m.s. ball speed, providing a separation of timescales for fast and slow variables, and implying that the drag and random forces should be independent of ball velocity. Such a separation of timescales is a necessary but not a sufficient condition for a Langevin description. Applicability of statistical mechanics also requires the complicated effects of turbulence to decouple into three distinct, additive components: (1) a force arising from a

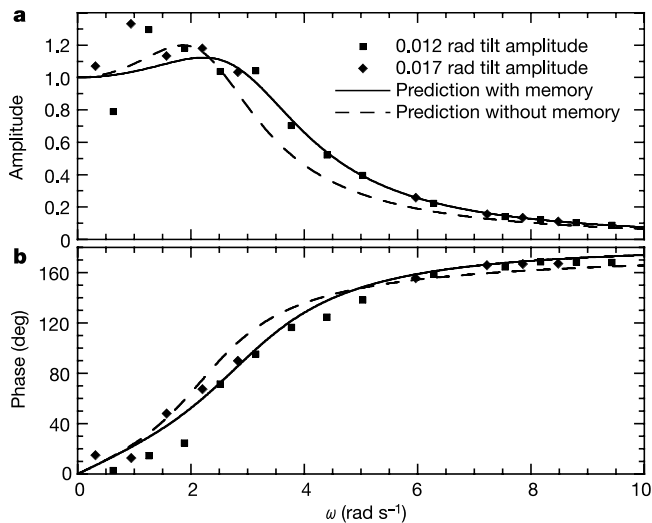


Figure 3 Dimensionless amplitude (a), and phase (b), of the average response of the sphere to sinusoidal tilting at frequency ω . The curves show expectations corresponding to the Langevin equation fits of Fig. 2.

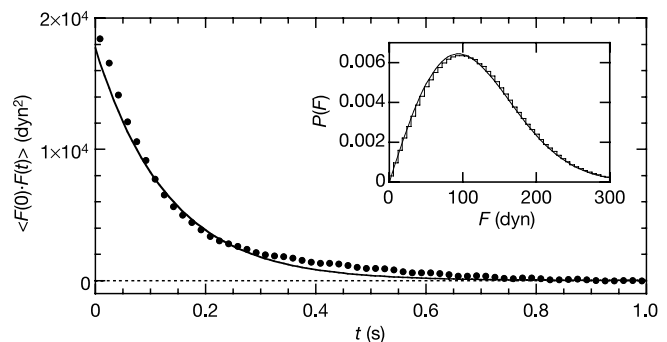


Figure 4 Autocorrelation and distribution (inset) of the random force acting on the sphere. The instantaneous random force is found from the time series data by subtracting the spring and drag forces from the instantaneous acceleration times m_e . The curves show expectations corresponding to the Langevin equation fit, with memory, of Fig. 2.

harmonic confining potential; (2) a position-independent, frequency-dependent drag; and (3) a position-independent, gaussian-distributed, temporally correlated random force. It is not a priori obvious that this decoupling should occur. The fact that it does could be significant for the aerodynamics of nonsteady forces on bluff bodies in turbulent flows—a field of research with many important applications²⁸.

This work provides an exhaustive demonstration of a steadily driven dissipative system that is perfectly described by near-equilibrium statistical mechanics. Unlike driven systems close to jamming, ours has a single effective temperature that holds at all timescales and that therefore completely describes the fluctuating dynamics. Given this result we can now alter the system, seeking to progressively violate the thermal analogy. Thus, gas-fluidized ping-pong balls could provide an experimental means of addressing the broader challenge of understanding the extent to which near-equilibrium statistical mechanics applies to driven systems. □

Methods

The sphere is a hollow ping-pong ball of mass $m = 2.55$ g, diameter $D = 3.8$ cm, and shell thickness 0.038 cm; including rolling inertial mass, its effective mass is $m_e = m + I/R^2 = 4.22$ g. It rolls without slipping on a flat 12-inch diameter brass sieve, with a wire mesh spacing of $300\ \mu\text{m}$, and with a 4-inch side wall. Variation of mesh size and height of side wall has little effect. The speed and uniformity of the upflow of air are monitored with a hot-wire anemometer. Here, we work at a flow speed of $U = 280\ \text{cm s}^{-1}$; this is below the terminal speed of a falling ping-pong ball ($800\ \text{cm s}^{-1}$), and gives a Reynolds number, based on sphere size, of $\text{Re} = 10^4$. Thus, the ball generates turbulence and this kicks it around in the plane. Uniformity of airflow, verified with the anemometer, is achieved by mounting the sieve to a 20 inch $\times$ 20 inch $\times$ 4 foot tall windbox consisting of two nearly cubical chambers separated by a perforated metal sheet. Air from a blower is introduced through a cloth sleeve connected to an input port on the side of the lower chamber. The windbox is mounted on a motorized tilting platform, which in turn is mounted on a rigid frame with a three-point levelling mechanism. The ball position is measured by a digital camera at a frame rate of 120 Hz. The ball speed and acceleration are found by fitting position versus time to a third-order polynomial, with a fitting window of ± 4 points and a gaussian weighting that is nearly zero at the edges. The position resolution is ± 0.05 mm. The data in Figs 1–4 are based on 3.9 h of position measurements.

Received 29 May; accepted 16 December 2003; doi:10.1038/nature02294.

- Hohenberg, P. C. & Shraiman, B. I. Chaotic behavior of an extended system. *Physica D* **37**, 109–115 (1989).
- Egolf, D. A. Equilibrium regained: From nonequilibrium chaos to statistical mechanics. *Science* **287**, 101–104 (2000).
- Cugliandolo, L. F. & Kurchan, J. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Phys. Rev. Lett.* **71**, 173–176 (1993).
- Herisson, D. & Ocio, M. Fluctuation-dissipation ratio of a spin glass in the aging regime. *Phys. Rev. Lett.* **88**, 257202 (2002).
- Grigera, T. S. & Israeloff, N. E. Observation of fluctuation-dissipation-theorem violations in a structural glass. *Phys. Rev. Lett.* **83**, 5038–5041 (1999).
- Berthier, L. & Barrat, J. L. Shearing a glassy material: numerical tests of nonequilibrium mode-coupling approaches and experimental proposals. *Phys. Rev. Lett.* **89**, 095702 (2002).
- Segre, P. N., Liu, F., Umbanhowar, P. B. & Weitz, D. A. An effective gravitational temperature for sedimentation. *Nature* **409**, 594–597 (2001).
- Bellon, L., Ciliberto, S. & Laroche, C. Violation of the fluctuation-dissipation relation during the formation of a colloidal glass. *Europhys. Lett.* **53**, 511–517 (2001).
- Mueh, D. M. *et al.* Signatures of granular microstructure in dense shear flows. *Nature* **406**, 385–389 (2000).
- Losert, W., Bocquet, L., Lubensky, T. C. & Gollub, J. P. Particle dynamics in sheared granular matter. *Phys. Rev. Lett.* **85**, 1428–1431 (2000).
- Reydellet, G., Rioual, F. & Clement, E. Granular hydrodynamics and density wave regimes in a vertical chute experiment. *Europhys. Lett.* **51**, 27–33 (2000).
- Lemieux, P. A. & Durian, D. J. From avalanches to fluid flow: a continuous picture of grain dynamics down a heap. *Phys. Rev. Lett.* **85**, 4273–4276 (2000).
- Makse, H. A. & Kurchan, J. Testing the thermodynamic approach to granular matter with a numerical model of a decisive experiment. *Nature* **415**, 614–617 (2002).
- D'Anna, G., Mayor, P., Barrat, A., Loreto, V. & Nori, F. Observing brownian motion in vibration-fluidized granular matter. *Nature* **424**, 909–912 (2003).
- Ono, I. K. *et al.* Effective temperatures of a driven system near jamming. *Phys. Rev. Lett.* **89**, 095703 (2002).
- Jaeger, H. M., Nagel, S. R. & Behringer, R. P. Granular solids, liquids, and gases. *Rev. Mod. Phys.* **68**, 1259–1273 (1996).
- Duran, J. *Sands and Powders, and Grains: An Introduction to the Physics of Granular Materials* (Springer, New York, 2000).
- Geldart, D. *Gas Fluidization Technology* (Wiley, New York, 1986).
- Poulligny, B., Malzbender, R., Ryan, P. & Clark, N. A. Analog simulation of melting in two dimensions. *Phys. Rev. B* **42**, 988–991 (1990).
- Ippolito, L., Annic, C., Lemaître, J., Oger, L. & Bideau, D. Granular temperature: experimental analysis. *Phys. Rev. E* **52**, 2072–2075 (1995).

- Olafsen, J. S. & Urbach, J. S. Clustering, order, and collapse in a driven granular monolayer. *Phys. Rev. Lett.* **81**, 4369–4372 (1998).
- Feitosa, K. & Menon, N. Breakdown of energy equipartition in a 2D binary vibrated granular gas. *Phys. Rev. Lett.* **88**, 198301 (2002).
- Baxter, G. W. & Olafsen, J. S. Gaussian statistics in granular gases. *Nature* **425**, 680 (2003).
- Prentis, J. J. Experiments in statistical mechanics. *Am. J. Phys.* **68**, 1073–1083 (2000).
- Kubo, R., Toda, M. & Hashitsume, N. *Statistical Physics II: Nonequilibrium Statistical Mechanics* (Springer, New York, 1991).
- Achenbach, E. Vortex shedding from spheres. *J. Fluid Mech.* **62**, 209–221 (1974).
- Suryanarayana, G. K. & Prabhu, A. Effect of natural ventilation on the boundary layer separation and near-wake vortex shedding characteristics of a sphere. *Exp. Fluids* **29**, 582–591 (2000).
- Leweke, T., Bearman, P. W. & Williamson, C. H. K. Special issue on bluff body wakes and vortex-induced vibrations—Preface. *J. Fluids Struct.* **15**, 377–378 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank L. Bocquet, R.F. Bruinsma, P.G. de Gennes, J. B. Freund, D. Levine, and M. A. Rutgers for suggestions. This work was supported by the NSF through grants to D.J.D. and A.J.L.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.D. (durian@physics.ucla.edu).

A route to high surface area, porosity and inclusion of large molecules in crystals

Hee K. Chae^{1*}, Diana Y. Siberio-Pérez^{1,2}, Jaheon Kim¹, YongBok Go¹, Mohamed Eddaoudi¹, Adam J. Matzger^{1,2}, Michael O'Keeffe³ & Omar M. Yaghi¹

Materials Design and Discovery Group

¹Department of Chemistry, ²Macromolecular Science and Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA

³Department of Chemistry, Arizona State University, Tempe, Arizona 85287, USA

* Permanent address: Department of Chemistry, Hankuk University of Foreign Studies, Korea

One of the outstanding challenges in the field of porous materials is the design and synthesis of chemical structures with exceptionally high surface areas¹. Such materials are of critical importance to many applications involving catalysis, separation and gas storage. The claim for the highest surface area of a disordered structure is for carbon, at $2,030\ \text{m}^2\ \text{g}^{-1}$ (ref. 2). Until recently, the largest surface area of an ordered structure was that of zeolite Y, recorded at $904\ \text{m}^2\ \text{g}^{-1}$ (ref. 3). But with the introduction of metal-organic framework materials, this has been exceeded, with values up to $3,000\ \text{m}^2\ \text{g}^{-1}$ (refs 4–7). Despite this, no method of determining the upper limit in surface area for a material has yet been found. Here we present a general strategy that has allowed us to realize a structure having by far the highest surface area reported to date. We report the design, synthesis and properties of crystalline $\text{Zn}_4\text{O}(1,3,5\text{-benzenetribenzoate})_2$, a new metal-organic framework with a surface area estimated at $4,500\ \text{m}^2\ \text{g}^{-1}$. This framework, which we name MOF-177, combines this exceptional level of surface area with an ordered structure that has extra-large pores capable of binding polycyclic organic guest molecules—attributes not previously combined in one material.

The conceptual basis of our strategy can be appreciated by considering a graphene sheet (Fig. 1a). Excision of progressively smaller fragments from this sheet and calculation of their Connolly surface areas⁸ (see Methods) shows that exposing the latent edges of

the six-membered rings leads to significant enhancement of specific surface area. Thus the surface area of a single infinite sheet is $2,965 \text{ m}^2 \text{ g}^{-1}$ (calculating both sides; see Methods). For units consisting of infinite chains of poly-*p*-linked six-membered rings (Fig. 1b), the surface area is almost doubled ($5,683 \text{ m}^2 \text{ g}^{-1}$).

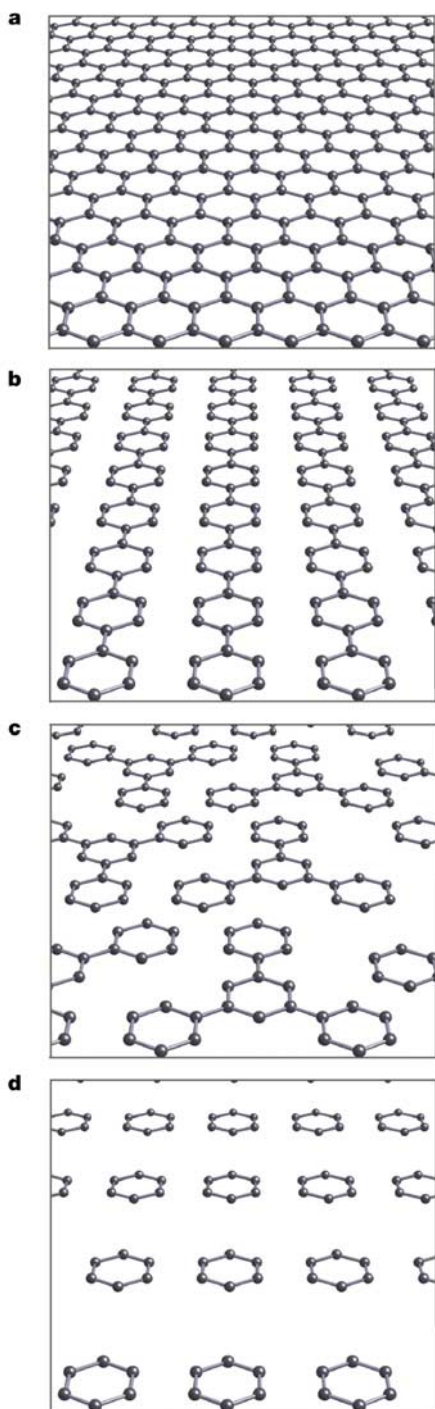


Figure 1 The surface area of graphite fragments. **a**, A graphene sheet extracted from the graphite structure has a Connolly surface area of $2,965 \text{ m}^2 \text{ g}^{-1}$ when calculated in Cerius². **b**, A series of poly-*p*-linked six-membered rings can be extracted from that sheet, thus increasing the surface area to $5,683 \text{ m}^2 \text{ g}^{-1}$. **c**, Excision of six-membered rings 1,3,5-linked to a central ring raises the surface area to $6,200 \text{ m}^2 \text{ g}^{-1}$. **d**, The surface area reaches a maximum of $7,745 \text{ m}^2 \text{ g}^{-1}$ when the graphene sheet is fully decomposed into isolated six-membered rings.

Alternatively, if the graphene sheet is divided into units of three six-membered rings that are 1,3,5-linked to a central ring (Fig. 1c), the surface area is similarly high ($6,200 \text{ m}^2 \text{ g}^{-1}$). Finally, exposing all latent edges to give isolated six-membered rings (Fig. 1d) leads to an upper limit value of $7,745 \text{ m}^2 \text{ g}^{-1}$. This analysis does not take into account the hydrogen atoms that would terminate the fragments in metal-organic frameworks (MOFs), although that would result in even higher surface areas for those fragments. Linking these fragments into a real material will not lead to complete realization of these theoretically limiting values. Nevertheless, this analysis suggests that structures with condensed rings should be avoided in order to maximize the number of exposed ring faces and edges.

To put these ideas into practice, we used established reticular chemistry⁹ reactions to link the carboxylate derivative (1,3,5-benzenetricarboxylate (BTB), a triangular unit of the type shown in Fig. 1c) with basic zinc(II) carboxylate clusters ($\text{Zn}_4\text{O}(\text{CO}_2)_6$, an octahedral unit, see Fig. 2a) into MOF-177. Block-shaped crystals of MOF-177 were produced (see Methods) by heating a mixture of H_3BTB and $\text{Zn}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ in *N,N*-diethylformamide (DEF) to 100°C . The crystals were formulated by elemental analysis as $\text{Zn}_4\text{O}(\text{BTB})_2 \cdot (\text{DEF})_{15}(\text{H}_2\text{O})_3$ (Calculated: C, 56.96; H, 7.46; N, 7.73. Found: C, 56.90; H, 7.54; N, 7.67). An X-ray diffraction study (see Methods and Supplementary Information) on a crystal isolated from the reaction mixture confirmed this formulation. It also revealed a remarkably open three-dimensional structure of composition $\text{Zn}_4\text{O}(\text{BTB})_2$, in which each basic zinc acetate cluster is linked to six BTB units (Fig. 2b). In this structure there are 84 exposed edges (60 C–C, 12 C–O, and 12 Zn–O) and only four fused edges (Zn–O) per formula unit (Fig. 2c). The structure of MOF-177 is entirely constructed of six-membered C_6H_4 , C_6H_3 and OZn_2CO_2 rings.

There are two places in the structure maximally far from any framework atom. Positions (0,0,0) and (0,0,1/2) have the nearest carbon atom at 7.6 Å and the six positions at (1/2,0,0), etc., have the nearest carbon atom at 7.1 Å. Allowing for a carbon atom van der Waals radius of 1.7 Å, these accommodate spheres of diameter 11.8 and 10.8 Å respectively without touching any framework atoms. The latter pores are connected to produce continuous sinuous channels along (1/2,0,z), (0,1/2,z), and (1/2,1/2,z) (see Fig. 2c). In the as-prepared material, the cavities are occupied by at least 15 DEF and 3 H_2O guests per formula unit. The space occupied by guests alone is 81% of the cell volume. Indeed, gas sorption studies indicate that this space is accessible to incoming guest species and that the framework maintains its integrity in the absence of guests.

Evidence of guest mobility and framework stability initially came from a thermal gravimetric analysis study. A crystalline sample (2.946 mg) was heated at a constant rate of 5°C min^{-1} in air from 25 to 600°C . Two weight-loss steps were observed: the first, corresponding to 47.95%, occurred between 50 and 100°C , which can be attributed to the loss of guest molecules (calculated 48.17%); the second weight loss of 22.01% above 350°C is due to decomposition of the framework. The lack of any weight loss between 100 and 350°C indicated that the framework is thermally stable in air at those temperatures (see Supplementary Information). Comparison of the X-ray powder diffraction patterns of the as-synthesized MOF-177 with samples of the material having completely evacuated pores show that the framework periodicity and structure are still preserved, further confirming the architectural stability of the framework in the absence of guests (see Supplementary Information).

To determine the capacity of this material for the uptake of gases, we measured the gaseous N_2 sorption isotherm on samples of MOF-177 in which the pores were fully evacuated. The isotherm revealed a reversible type I behaviour and showed no hysteresis upon desorption of gas from the pores (Fig. 3). The accessible void space is fully saturated with N_2 molecules at relatively low pressures ($P/P_0 \approx 0.2$) with a total weight uptake of 1,288 mg N_2 per gram of

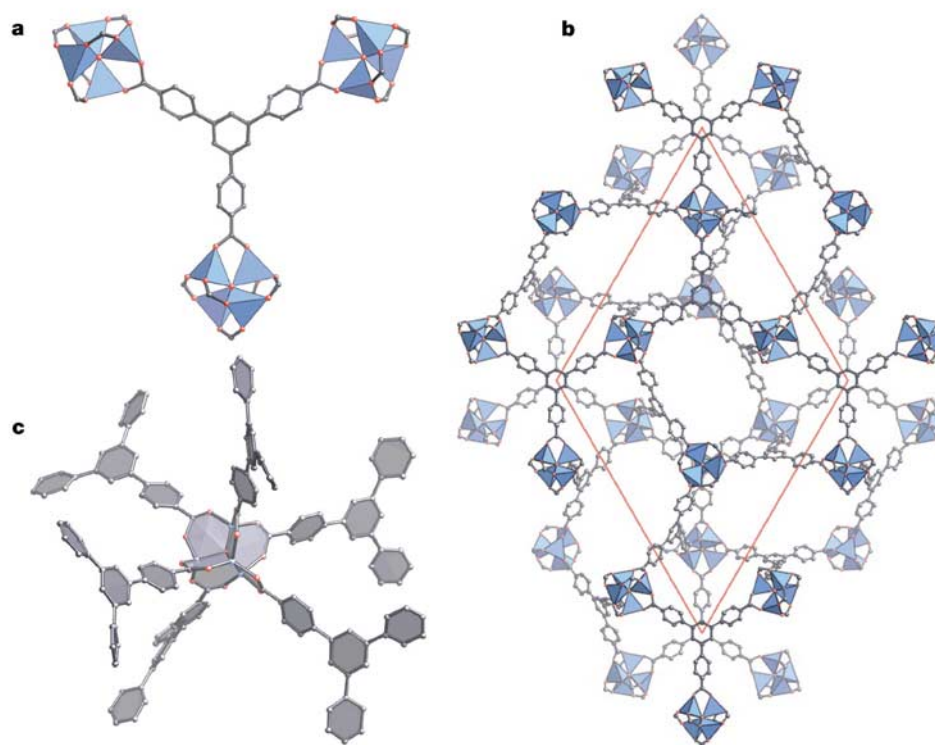


Figure 2 The structure of MOF-177. **a**, A BTB unit linked to three OZn₄ units (H atoms are omitted). ZnO₄ tetrahedra are shown in blue and O and C atoms are shown as red and black spheres, respectively. **b**, The structure projected down [001]. Colours as for **a**. For

clarity only about half the *c*-axis repeat unit is shown. **c**, A fragment of the structure radiating from a central OZn₄; six-membered rings are shown as grey hexagons and Zn atoms as blue spheres.

the fully evacuated framework, which correlates to an estimated total number of N₂ molecules of 52.7 per formula unit, and 422 per unit cell.

Using the Dubinin–Raduskhvich equation, we obtained a pore volume of 1.59 cm³ g⁻¹ (0.69 cm³ cm⁻³). By assuming a monolayer coverage of N₂ and applying the Langmuir model, we find the apparent Langmuir surface area to be 4,500 m² g⁻¹. The narrowest dimension (10.8 Å) of the pores of MOF-177 is still in the microporous regime (<20 Å diameter pore size). The absolute value of the surface area of all such materials is subject to systematic error arising from standard considerations involving (1) the area assigned to the absorbed molecules and (2) the assumption of monolayer coverage. However, plateaux in isotherms of the type shown in Fig. 3 are invariably interpreted as corresponding to monolayer coverage, so values for different materials, obtained by different groups, are comparable. Nevertheless, the pore volume and surface area of MOF-177 are well beyond those observed for the most porous crystalline zeolites and porous carbon and significantly exceed the previous record^{4,9} for a crystalline MOF material (0.59 cm³ cm⁻³ and 2,900 m² g⁻¹ for MOF-5).

The underlying topology of MOF-177 is a (6,3)-coordinated net with the centre of the octahedral OZn₄(CO₂)₆ cluster as the site of six-coordination and the centre of the BTB unit the site of three-coordination. The structure of this net plays an important part in determining pore size by preventing the formation of interpenetrating frameworks. The most regular ('default') (6,3)-coordinated net is 'pyr', named after the pyrite structure¹⁰. However, two such nets can interpenetrate in such a way that all the rings of one structure are penetrated by the links of the other ('fully catenated') and vice versa, and indeed MOF-150, on the basis of this topology, occurs as an interpenetrating pair of nets (Fig. 4a)¹¹. The second net that fully catenates a given net is said to be the dual of that net, and if

a net and its dual have the same structure (as in the case of pyr) they are said to be self-dual. Although self-duality is a rare property of nets, it does occur also for default structures of nets with three-, four- and six-coordination, and thus interpenetration of two (or more) copies of identical nets is found to be a common obstacle to synthesis of large-pore materials.

We show here that an effective strategy for avoiding interpenetration is to use nets for which the structure of the dual is very different. The net underlying MOF-177 (Fig. 2b), which we term 'qom', is related to the pyr net. In pyr the six-coordinated sites are arranged as the centres of the spheres in cubic closest packing (that is, on a face-centred cubic lattice); in qom the corresponding

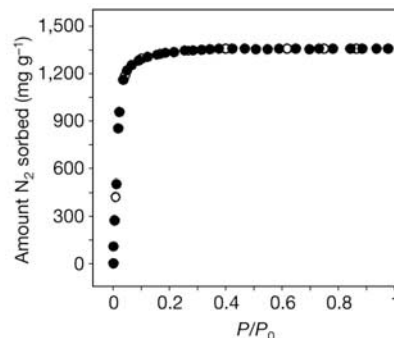


Figure 3 Nitrogen gas sorption isotherm at 78 K for MOF-177 (filled circles, sorption; open circles, desorption). P/P_0 is the ratio of gas pressure (P) to saturation pressure (P_0), with $P_0 = 746$ torr.

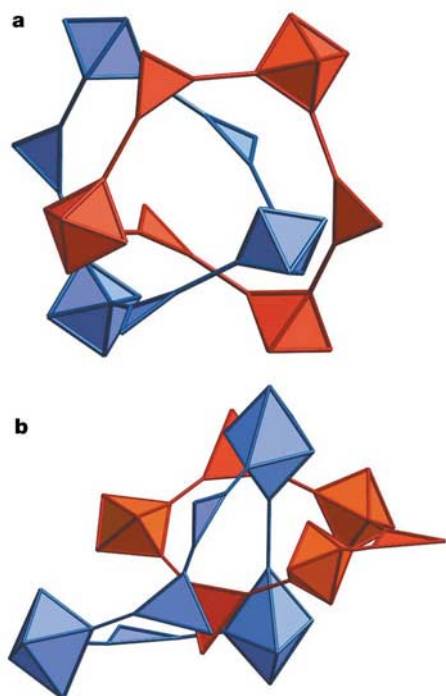


Figure 4 Catenation of rings in nets intergrown with their dual structures. The nets are shown augmented with triangles at the three-coordinated vertices and octahedra at the six-coordinated vertices. **a**, A pair of identical rings in the self-dual pyr net of MOF-150 (ref. 11). **b**, A six-membered ring of the qom net (red) of MOF-177 catenated with a ring of the dual net (blue). A pair of three-coordinated vertices are directly linked, as are pairs of six-coordinated vertices.

arrangement is that of hexagonal closest packing. However, the dual net, although also (6,3)-coordinated, is very different, and as some of the edges link sites of the same coordination (Fig. 4b) it is not a viable candidate for a MOF. Likewise, as qom is very different from its dual, two such qom nets cannot interpenetrate efficiently. First, we need to identify the strategy for avoiding the pyr net. Simple geometrical arguments show that to prevent formation of the pyr net (as found in MOF-150), when linking octahedral $\text{OZn}_4(\text{CO}_2)_6$, one should employ aromatic tricarboxylates such as BTB, which is known to have coplanar carboxylates in MOFs¹².

Given the exceptional stability, porosity and large pores of MOF-177, we sought to test its ability to adsorb large organic molecules. Traditionally, inclusion in porous materials has been achieved by either *in situ* synthesis of the guest, synthesis of the framework to entrap the guest or direct incorporation by adsorption¹³. The former two methods are not well suited to making new materials for separations. Furthermore, in all three methods the use of polycrystalline materials raises the concern that inclusion takes place in intercrystalline regions rather than directly in the pores¹⁴. We have circumvented this concern by using monocrystalline samples of MOF-177 in all studies. Initial studies demonstrated facile uptake of bromobenzene, 1-bromonaphthalene, 2-bromonaphthalene and 9-bromoanthracene from solution (see Methods). However, the uniformity of distribution of these guests in the crystals was difficult to determine directly. Thus we included coloured organic molecules in MOF-177 single crystals so that incorporation of the guest could be directly verified visually¹⁵.

MOF-177 crystals were placed in a C_{60} -toluene solution. After several days the crystals' shape and integrity remained intact and a change in colour to deep red provided optical evidence of C_{60}

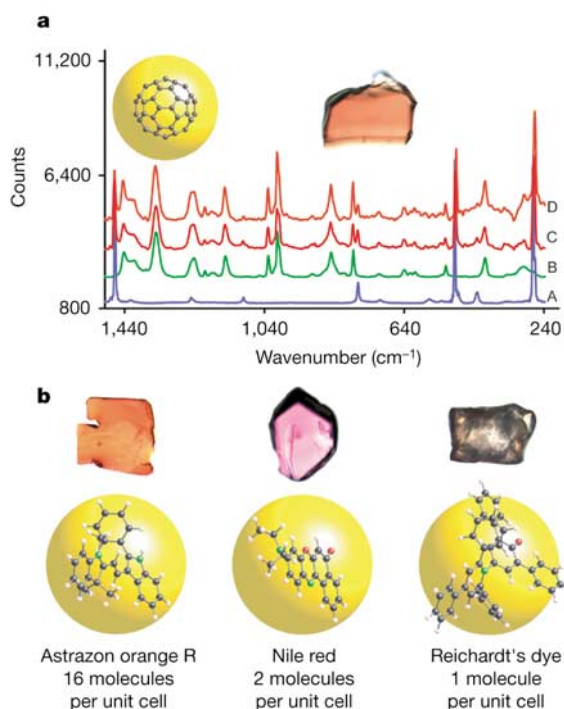


Figure 5 Inclusion of polycyclic organic guests. **a**, Colourless crystals turned deep red, indicating adsorption of C_{60} into MOF-177 single crystals. Analytical evidence was provided by comparison of Raman spectra of a sliced crystal (D) and a whole crystal (C) with bulk C_{60} (A) and an evacuated MOF (B). **b**, The ability of MOF-177 crystals to adsorb large guests was quantified for the dyes Astrazon Orange R, Nile Red and Reichardt's dye. These incorporated 16, 2 and 1 molecules per unit cell, respectively. The ball-and-stoke drawings of the molecules are superimposed on a ball of 11 Å diameter that fits into the pores of MOF-177.

inclusion in the framework (Fig. 5a). To probe the presence of C_{60} , a MOF-177- C_{60} complex was analysed by Raman spectroscopy. This vibrational spectrum was compared with spectra of bulk C_{60} and with that of evacuated MOF-177. The encapsulated fullerene complex exhibited bands at the same positions as the desolvated MOF-177. However, the fullerene bands were broadened and observed at positions slightly shifted from bulk C_{60} , indicating interaction with the framework (Fig. 5a). Uniformity of inclusion was assessed by slicing a single crystal into three parts, thus exposing the inner core, and verifying that the middle portion was evenly coloured throughout and that the Raman spectrum exhibited bands for both framework and guest (Fig. 5a).

To quantify the ability of MOF-177 to accommodate large polycyclic organic molecules, three dyes, Astrazon Orange R, Nile Red and Reichardt's dye, were selected. We used saturated solutions of these compounds to dye the crystals, and examined a section from the centre of the MOF-177 crystal to gauge the uniformity of dye distribution (Fig. 5b; see Methods). Astrazon Orange R and Nile Red coloured the slices uniformly, indicating free movement of the dye into the crystals. Astrazon Orange R achieved over 40 wt% in the crystals, corresponding to 16 dye molecules in each unit cell. On average, two Nile Red molecules entered each unit cell. The very large molecules of Reichardt's dye, however, penetrated only the outer part of the crystal, with only 1 molecule entering each unit cell on average. Together with the diffusion experiments, these results clearly demonstrate the potential for size selectivity in a regime that is currently inaccessible with conventional porous materials. □

Methods

Surface area calculations

The surface areas for graphite, and fragments of this structure were obtained via the Connolly Surface method⁸, as implemented by Cerius² 4.2 from Accelrys.

Synthesis of MOF-177

A solution of DEF containing 4,4',4''-benzene-1,3,5-triyl-tri-benzoic acid (H₃BTB; 5.00×10^{-3} g, 1.14×10^{-5} mol) and Zn(NO₃)₂·6H₂O (0.020 g, 6.72×10^{-5} mol) was placed in a Pyrex tube of dimensions 10 mm (outer diameter), 8 mm (inner diameter) and 150 mm (length). The sealed tube was heated at a rate of $2.0^\circ\text{C min}^{-1}$ to 100°C , held at 100°C for 23 h, and cooled at a rate of $0.2^\circ\text{C min}^{-1}$ to room temperature. Block-shaped crystals of MOF-177 were formed and isolated by washing with DEF (4 × 2 ml) and drying briefly in air (~1 min) (0.005 g, 32% based on H₃BTB).

Crystallographic studies on MOF-177

Crystal (0.30 × 0.30 × 0.28 mm³) of Zn₄O(BTB)₂·(DEF)₁₅·(H₂O)₃ was sealed in a glass capillary and mounted on a Bruker SMART APEX charge-coupled device diffractometer equipped with a normal focus Mo-target X-ray tube ($\lambda = 0.71073 \text{ \AA}$) operated at 2,000 W power (50 kV, 40 mA). The X-ray intensities were measured at 273(2) K. A total of 1,800 frames were collected with a scan width of 0.3° in ω with an exposure time of 30 s per frame. The frames were integrated with the SAINT software package with a narrow frame algorithm. The integration of the data using a trigonal unit cell yielded a total of 173,392 reflections to a maximum 2θ value of 41.68° of which 12,530 were independent and 5,233 were greater than $2\sigma(I)$. The final cell constants were refined with 5,049 reflections with $4.395 < 2\theta < 41.661$. Analysis of the data showed negligible decay during data collection. Absorption correction was applied by using SADABS. The structure was solved by direct methods and the subsequent difference Fourier syntheses and refined with the SHELXTL (version 6.10) software package, using the trigonal space group $P\bar{3}1c$ (number 163), $a = 37.072(2) \text{ \AA}$, $c = 30.033(2) \text{ \AA}$ with $Z = 8$ for the formula based on the elemental analysis. There were two independent Zn₄O clusters centred at Wyckoff positions $2d$ and $6h$: the first of these was disordered over two possible orientations. Final full-matrix least-squares refinement on F^2 converged to $R1 = 0.1538$ ($F > 4\sigma(F)$) and $wR2 = 0.4639$ (all data) with GOF = 1.397. Additional details are presented as Supplementary Information.

Diffusion of bromoarenes

The crystals were transferred from their mother liquor (DMF) to heptane. After 30 min, the heptane was removed and fresh heptane was once again added. This process was repeated three times in order to ensure complete displacement of DMF molecules from the porous framework. The excess heptane was then removed and 1 ml of a heptane solution containing 0.007 M of each of the four bromoarenes was added. The crystals remained immersed in this solution for 90 min. The concentration of each bromoarene in the supernatant liquid was monitored by gas chromatography. The disappearance of material indicates adsorption of bromoarenes by MOF-177 crystals.

Quantification of dye uptake

MOF-177 crystals (3–5 mg) were placed in 0.15 ml of a saturated solution of dye in CH₂Cl₂. During a period of 6 days, the supernatant solution was removed and replaced with fresh dye solution 20 times. After the sixth day of inclusion, the crystals were removed from solution and rinsed three times with CH₂Cl₂. Individual crystals were precisely weighed with a microgram balance and digested in 40 to 60 μl of 0.1 M NaOH in methanol. This solution was quantitatively transferred to a 2 ml volumetric flask and methanol was added to obtain precise dilution. Ultraviolet–visual spectrum absorbance analysis of the resulting solutions allowed for determination of the concentrations of the dyes and thus for the amount of dye included in the MOF-177 framework.

Received 9 September; accepted 16 December 2003; doi:10.1038/nature02311.

1. Davis, M. E. Ordered porous materials for emerging applications. *Nature* **417**, 813–821 (2002).
2. Nijkamp, M. G., Raaymakers, J. E., van Dillen, A. J. & de Jong, K. P. Hydrogen storage using physisorption-materials demands. *Appl. Phys. A* **72**, 619–623 (2001).
3. Chester, A. W., Clement, P. & Han, S. Faujasite zeolitic materials. US patent 6,136,291A (24 October 2000).
4. Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **402**, 276–279 (1999).
5. Eddaoudi, M. *et al.* Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science* **295**, 469–472 (2002).
6. Noro, S. *et al.* Framework engineering by anions and porous functionalities of Cu(II)/4,4'-bpy coordination polymers. *J. Am. Chem. Soc.* **2002**, 2568–2583 (2002).
7. Seki, K., Takamizawa, S. & Mori, W. Design and gas adsorption property of a three-dimensional coordination polymer with a stable and highly porous framework. *Chem. Lett.* **30**, 332–333 (2001).
8. Connolly, M. L. Solvent accessible surfaces of proteins and nucleic acids. *Science* **221**, 709–713 (1983).
9. Yaghi, O. M. *et al.* Reticular synthesis and the design of new materials. *Nature* **423**, 705–714 (2003).
10. Delgado-Friedrichs, O., O'Keeffe, M. & Yaghi, O. M. Three-periodic nets and tilings: regular and quasiregular nets. *Acta Crystallogr. A* **59**, 22–27 (2003).
11. Chae, H. K., Kim, J., Delgado-Friedrichs, O., O'Keeffe, M. & Yaghi, O. M. Design of frameworks with mixed triangular and octahedral building blocks exemplified by the structure of Zn₄O(TCA)₂ having the pyrite topology (TCA = 4,4',4'' tricarboxytriphenylamine). *Angew. Chem. Int. Edn Engl.* **42**, 3907–3909 (2003).
12. Chen, B., Eddaoudi, M., Hyde, T., O'Keeffe, M. & Yaghi, O. Interwoven metal-organic framework on a periodic minimal surface with extra-large pores. *Science* **291**, 1021–1023 (2001).

13. De Vos, D. E. & Jacobs, P. A. Zeolite-based supramolecular assemblies. *Stud. Surf. Sci. Catal.* **137**, 957–985 (2001).
14. Sato, K., Nishimura, Y., Honna, K., Matsubayashi, N. & Shimada, H. Role of HY zeolite mesopores in hydrocracking of heavy oils. *J. Catal.* **200**, 288–297 (2001).
15. Kahr, B. & Gurney, R. W. Dyeing crystals. *Chem. Rev.* **101**, 893–951 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Initial phases of this work were carried out by H. Li and scale-up was performed by A. Benin. We are grateful to the NSF and the DOE for support of various aspects of this programme.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A. J. M. (matzger@umich.edu), M.O.K. (mokeeffe@asu.edu) or O.M.Y. (oyaghi@umich.edu).

Hydrogenation and cleavage of dinitrogen to ammonia with a zirconium complex

Jaime A. Pool, Emil Lobkovsky & Paul J. Chirik

Department of Chemistry and Chemical Biology, Baker Laboratory Cornell University, Ithaca, New York 14853, USA

Molecular nitrogen is relatively inert owing to the strength of its triple bond, nonpolarity and high ionization potential. As a result, the fixation of atmospheric nitrogen to ammonia under mild conditions has remained a challenge to chemists for more than a century. Although the Haber–Bosch process produces over 100 million tons of ammonia annually¹ for the chemical industry and agriculture², it requires high temperature and pressure, in addition to a catalyst³, to induce the combination of hydrogen (H₂) and nitrogen (N₂). Coordination of molecular nitrogen to transition metal complexes can activate and even rupture the strong N–N bond⁴ under mild conditions, with protonation yielding ammonia in stoichiometric⁵ and even catalytic yields⁶. But the assembly of N–H bonds directly from H₂ and N₂ remains challenging: adding H₂ to a metal–N₂ complex results in the formation of N₂ and metal–hydrogen bonds or, in the case of one zirconium complex⁷, in formation of one N–H bond and a bridging hydride. Here we extend our work on zirconium complexes containing cyclopentadienyl ligands^{8,9} and show that adjustment of the ligands allows direct observation of N–H bond formation from N₂ and H₂. Subsequent warming of the complex cleaves the N–N bond at 45°C , and continued hydrogenation at 85°C results in complete fixation to ammonia.

Coordination of N₂ to a homogeneous transition metal complex is often an effective strategy for activation of the N–N bond, although few well-defined metal–N₂ complexes are capable of nitrogen fixation⁴. Molybdenum and tungsten N₂ complexes, for example, can be protonated to yield ammonia⁵, and recently this approach has been made catalytic with activities rivaling the nitrogenase family of enzymes⁶.

Direct observation of N₂ hydrogenation with a homogeneous transition-metal complex under mild conditions has thus far remained elusive. Typically, addition of H₂ to a metal–N₂ complex results in a metal–hydrogen bond with expulsion of free N₂ (ref. 4). One exception is the dinuclear N₂ complex, [(P₂N₂)Zr]₂($\mu_2\eta^2, \eta^2$ -N₂) (where P₂N₂ = PhP(CH₂SiMe₂NSiMe₂CH₂)₂PPh, where Ph = phenyl), which reacts over the course of one week with 1–4 atmospheres of H₂ to form one N–H bond and a bridging

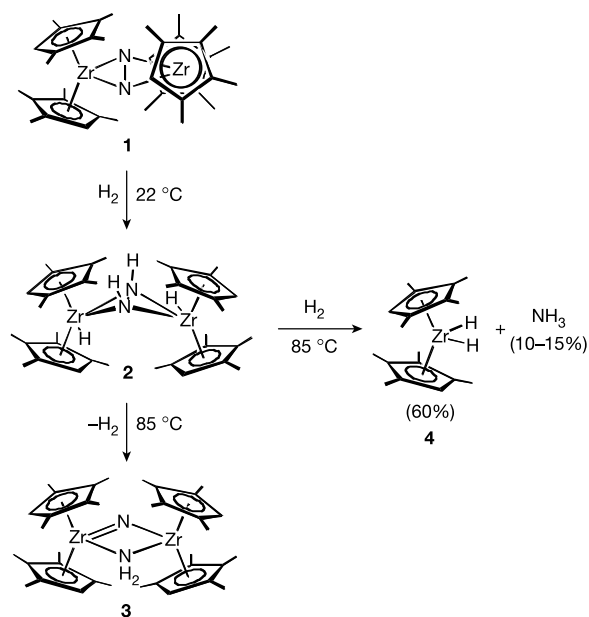


Figure 1 Hydrogenation and cleavage of N₂.

zirconium hydride⁷. Although computational studies on model compounds predict the reaction with additional H₂ to be thermodynamically feasible^{10,11}, further hydrogenation is not observed experimentally.

We have studied low-valent group 4 transition-metal complexes bearing cyclopentadienyl ligands (metallocenes) to understand their reactivity towards small molecules such as N₂ (ref. 8) and elemental phosphorus⁹. Specifically, we examined the influence of cyclopentadienyl ligand substitution on the hapticity of the N₂ ligand and how the coordination mode ultimately translates onto N₂ functionalization. As part of these investigations, systematic variation of the cyclopentadienyl methyl substituents on a low-valent zirconocene seemed a logical method for exploring structure–reactivity relationships of end-on versus side-on coordination.

Reduction of (η⁵-C₅Me₄H)₂ZrCl₂ (ref. 12) with sodium amalgam under one atmosphere of N₂ produced [(η⁵-C₅Me₄H)₂Zr]₂(μ₂,η²,η²-N₂) (structure 1) as a green solid that forms purple solutions when dissolved (Fig. 1). The solid-state structure of 1 is shown in Fig. 2 and reveals a side-on bound N₂ ligand contained in a planar Zr₂N₂ core. The cyclopentadienyl ligands on each zirconium are canted with respect to each other, forming a dihedral angle of 65.3°. The N–N bond length of 1.377(3) Å is elongated substantially compared to free N₂ (1.0975 Å) and the related end-on bound N₂ complex [(η⁵-C₅Me₅)₂Zr(η¹-N₂)]₂(μ₂,η¹,η¹-N₂) (ref. 13). However, this distance is shorter than that in hydrazine (1.460 Å) and other side-on bound zirconium N₂ complexes^{4,10,14}. In addition to X-ray diffraction, 1 was also characterized by multinuclear nuclear magnetic resonance (NMR) spectroscopy. The isotopically labelled complex 1-¹⁵N₂, [(η⁵-C₅Me₄H)₂Zr]₂(μ₂,η²,η²-¹⁵N₂), was prepared by addition of ¹⁵N₂ to 1 and displays a single resonance centred at 621.1 p.p.m. in the ¹⁵N NMR spectrum. This value is shifted slightly downfield relative to the value of 564 p.p.m. reported¹⁵ for [(η⁵-C₅Me₅)₂Zr(η¹-N₂)]₂(μ₂,η¹,η¹-N₂).

Addition of one atmosphere of H₂ gas to a pentane solution of 1 at ambient temperature resulted in immediate formation of a white precipitate identified as [(η⁵-C₅Me₄H)₂ZrH]₂(μ₂,η²,η²-N₂H₂) (structure 2) (Fig. 1). The ¹H NMR spectrum of 2 in benzene-*d*₆ displays singlets at 0.83 and 4.51 p.p.m., assignable to the N–H and Zr–H resonances, respectively. These peaks are absent in the ¹H

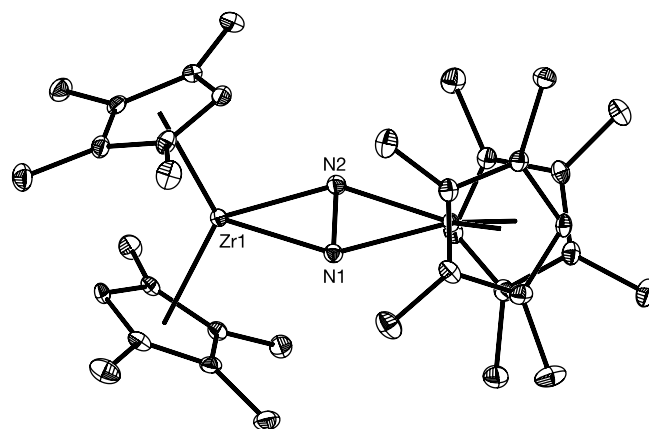


Figure 2 X-ray crystal structure of [(η⁵-C₅Me₄H)₂Zr]₂(μ₂,η²,η²-N₂) (1). Molecular structure at 30% probability ellipsoids, with hydrogen atoms omitted for clarity. Selected bond lengths and angles: N(1)–N(2): 1.377(3) Å, Zr(1)–N(1): 2.118(1) Å, Zr(1)–N(2): 2.131(1) Å, Zr(1A)–N(1): 2.119(1) Å, Zr(1A)–N(2): 2.131(1) Å, N(1)–Zr(1)–N(2): 37.81(9)°, N(2)–N(1)–Zr(1A): 171.57(7)°, N(2)–N(1)–Zr(1): 71.56(7)°, Zr(1A)–N(1)–Zr(1): 143.13(14)°, N(1)–N(2)–Zr(1A): 70.62(7)°, N(1)–N(2)–Zr(1): 70.62(7)°, Zr(1A)–N(2)–Zr(1): 141.25(14)°.

NMR spectrum upon addition of D₂ to 1 but may be observed by corresponding ²H NMR spectroscopy. For the hydrogenation of 1-¹⁵N₂, the peak at 0.83 p.p.m. splits into a two-line pattern with a coupling constant of 59 Hz, confirming its assignment as an N–H resonance. It should be noted that whereas an [AX]₂ pattern is theoretically possible for this isotopomer, only two broad lines are observed. A single peak centred at 74.5 p.p.m. is also observed in the ¹⁵N NMR spectrum in benzene-*d*₆ solution. Further confirmation of the hydrogenation of the N₂ ligand is provided by solution infrared spectroscopy, which exhibits absorptions at 3,235 and 1,554 cm^{−1} typical for N–H and Zr–H stretches, respectively. These peaks shift appropriately to 2,310 and 1,117 cm^{−1} upon deuteration.

The solid state structure of 2 was determined by single-crystal X-ray diffraction (Fig. 3). The crystal structure reveals a butterfly Zr(μ₂,η²,η²-N₂H₂)Zr core with a side-on bound diazenido ligand with an N–N bond length of 1.457(3) Å. The data were of sufficient quality that the hydrogens bound to the nitrogen atoms and the zirconium hydrides were located and freely refined. The N–H bonds are oriented perpendicular to the zirconocene wedges and the metal hydrides are arranged in a *transoid* fashion. Thus, both solution and solid-state characterization data clearly establish 2 as the first example of N₂ hydrogenation to hydrazine in a well-defined transition-metal complex. Significantly, the hydrogenation reaction is rapid and takes place with only one atmosphere of H₂ at ambient temperature. This behaviour contrasts with the reactivity of [(η⁵-C₅Me₅)₂Zr(η¹-N₂)]₂(μ₂,η¹,η¹-N₂) and other N₂ compounds which rapidly lose N₂ upon addition of H₂ (ref. 16). Although complete mechanistic details have not yet been elucidated, we believe that the unique reactivity of 1 may be attributed to significant zirconium-imido (Zr=NR)¹⁷ character of the N₂ ligand, possibly arising from facile side-on–end-on interconversion.

Heating solutions of 2 in boiling heptane (~100 °C) for 5 min resulted in loss of one equivalent of H₂ and cleavage of the N–N bond to form structure 3 [(η⁵-C₅Me₄H)₂Zr]₂(μ₂-NH₂)(μ₂-N) (Fig. 1). Monitoring of the reaction *in situ* by ¹H NMR spectroscopy revealed that N₂ cleavage can be induced at temperatures as low as 45 °C. However, for complete N₂ scission to occur, the liberated H₂ must be removed from the reaction mixture in order to drive the equilibrium.

Orange, benzene-*d*₆ solutions of 3 display four inequivalent

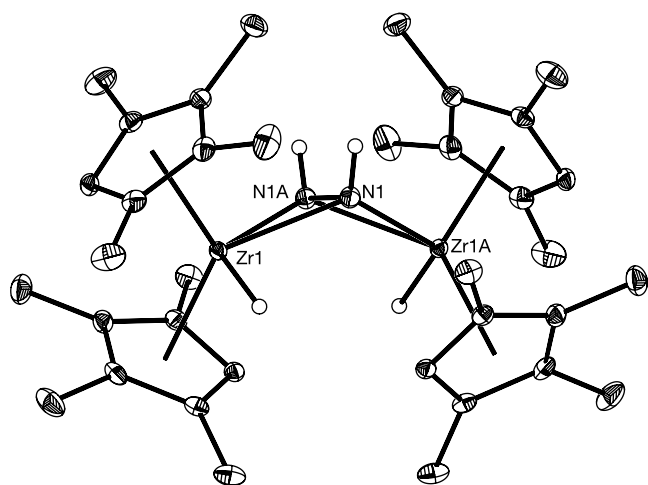


Figure 3 X-ray crystal structure of $[(\eta^5\text{-C}_5\text{Me}_4\text{H})_2\text{ZrH}]_2\text{N}_2\text{H}_2$ (**2**). Molecular structure at 30% probability ellipsoids, with hydrogen atoms on the cyclopentadienyl ligands omitted for clarity. Selected bond lengths and angles: Zr(1)–N(1A): 2.1974(13) Å, Zr(1)–N(1): 2.3163(14) Å, N(1)–N(1A): 1.457(3) Å, N(1)–Zr(1A): 2.1974(13) Å, N(1A)–Zr(1)–N(1): 37.55(6)°, N(1A)–N(1)–Zr(1A): 75.66(9)°, N(1A)–N(1)–Zr(1): 66.80(9)°, Zr(1A)–N(1)–Zr(1): 120.99(6)°.

cyclopentadienyl methyl groups by ^1H and ^{13}C NMR spectroscopy, consistent with C_s molecular symmetry arising from two different nitrogen environments in the zirconocene wedge. A single N–H resonance is observed at 4.58 p.p.m. by ^1H NMR spectroscopy and was also detected as a broad absorption centred at $3,681\text{ cm}^{-1}$ by solution infrared spectroscopy. Likewise, the ^{15}N NMR spectrum of isotopically enriched **3** in benzene- d_6 displays two resonances; one centred at 624.2 p.p.m. assigned to the bridging nitrido ligand and the other centred at 304.6 p.p.m. attributed to the bridging amido ligand.

The remaining challenge for nitrogen fixation was to form the additional N–H bonds and dissociate the ammonia from the zirconium centres. Treatment of **3** with excess anhydrous hydrochloric acid did indeed complete N_2 fixation, yielding two equivalents of NH_4Cl and $(\eta^5\text{-C}_5\text{Me}_4\text{H})_2\text{ZrCl}_2$. Anhydrous ammonia has been obtained by addition of stoichiometric quantities of acid and was detected by ^1H NMR spectroscopy, gas chromatography and indophenolic titration¹⁸. Thus, a combination of hydrogenolysis and protonolysis resulted in the complete fixation of N_2 to ammonia through a well-defined zirconium complex. Perhaps more significant is the first direct observation of N–H bond formation coupled with N_2 cleavage.

A more intriguing possibility is to achieve nitrogen fixation with H_2 as the sole source of N–H bonds. Gentle warming of heptane solutions of **2** to 85°C under one atmosphere of H_2 resulted in formation of ammonia, albeit in low (10–15%) yield, along with the dihydride, $(\eta^5\text{-C}_5\text{Me}_4\text{H})_2\text{ZrH}_2$ (structure **4**), and other zirconium(IV) compounds. Integration of the ^1H NMR spectrum indicates that approximately 60% of the total zirconium is converted to **4**. The ammonia was detected and quantified using a combination of NMR spectroscopy, gas chromatography and indophenolic titration¹⁸. Thus, nitrogen fixation has been accomplished under mild conditions using H_2 and a well-defined zirconium complex.

Although all manipulations were carried out using the most stringent anhydrous conditions, several control experiments were conducted to rule out ammonia generation by protonation from adventitious water. Reaction of the N_2 complex, **1**, or the dihydride, **4**, with trace amounts of water resulted in formation of the

zirconocene μ -oxo hydride complex, $[(\eta^5\text{-C}_5\text{Me}_4\text{H})_2\text{ZrH}]_2(\mu\text{-O})_2$ along with hydrazine and H_2 , respectively¹⁹. Addition of small (~10%) quantities of ammonia to **4** under an atmosphere of H_2 produced no reaction, although formation of the zirconocene amido hydride, $(\eta^5\text{-C}_5\text{Me}_4\text{H})_2\text{Zr}(\text{NH}_2)\text{H}$ (structure **5**), is observed when an excess of NH_3 is used. Treatment of **5** with an excess of water produced ammonia and the zirconocene bis-hydroxide complex, $(\eta^5\text{-C}_5\text{Me}_4\text{H})_2\text{Zr}(\text{OH})_2$. Both the μ -oxo hydride and the bis-hydroxide complex were not observed in the zirconocene product mixture after the hydrogenation of **2**, consistent with H_2 being the source of all three N–H bonds in the observed ammonia.

Insight into the origin of the ammonia was obtained by reaction of **2** with the dihydride, **4**. Immediately upon mixing, NH_3 is observed along with a complicated mixture of zirconocene products. Interestingly, these products are the same as those obtained after the hydrogenation of **2** at 85°C . One such compound that has been definitively characterized is the bridging imido-hydride complex, $[(\eta^5\text{-C}_5\text{Me}_4\text{H})_2\text{ZrH}]_2(\mu_2\text{-NH})$, which can also be independently prepared from reaction of the dihydride, **4**, with $(\eta^5\text{-C}_5\text{Me}_4\text{H})_2\text{Zr}(\text{NH}_2)\text{H}$. Determination of the identity of the zirconocene byproducts arising from this complicated series of transformations is currently under investigation. We expect that our direct observation of the key intermediates in the functionalization and cleavage pathway may aid the design of new stoichiometric and catalytic nitrogen fixation processes. □

Received 1 October; accepted 9 December 2003; doi:10.1038/nature02274.

- Schlögl, R. Catalytic synthesis of ammonia—A never-ending story. *Angew. Chem. Int. Edn Engl.* **42**, 2004–2008 (2003).
- Smil, V. *Enriching the Earth* (MIT Press, Cambridge, MA, 2001).
- Bielawa, H., Hinrichsen, O., Birkner, A. & Muhler, M. The ammonia-synthesis catalyst of the next generation: barium-promoted oxide-supported ruthenium. *Angew. Chem. Int. Edn Engl.* **40**, 1061–1063 (1999).
- Fryzuk, M. D. & Johnson, S. A. The continuing story of dinitrogen activation. *Coord. Chem. Rev.* **200**, 379–409 (2000).
- Hidai, M. Chemical nitrogen fixation by molybdenum and tungsten complexes. *Coord. Chem. Rev.* **186**, 99–108 (1999).
- Yandulov, D. M. & Schrock, R. R. Catalytic reduction of dinitrogen to ammonia at a single molybdenum center. *Science* **301**, 76–78 (2003).
- Fryzuk, M. D., Love, J. B. & Rettig, S. J. Transformation of coordinated dinitrogen by reaction with dihydrogen and primary silanes. *Science* **275**, 1445–1447 (1997).
- Pool, J. A., Lobkovsky, E. & Chirik, P. J. Cyclopentadienyl substituent effects on reductive elimination reactions in group 4 metallocenes: kinetics, mechanism and application to dinitrogen activation. *J. Am. Chem. Soc.* **125**, 2241–2251 (2003).
- Pool, J. A., Lobkovsky, E. & Chirik, P. J. Functionalization of elemental phosphorus with $[(\eta^5\text{-C}_5\text{Me}_5)(\eta^5\text{-C}_5\text{H}_3\text{Bu})\text{ZrH}_2]_2$. *Angew. Chem. Int. Edn Engl.* **41**, 3463–3466 (2002).
- Basch, H., Musaev, D. G. & Morokuma, K. Why does the reaction of the dihydrogen molecule with $[\text{P}_2\text{N}_2]\text{Zr}(\mu\text{-}\eta^2\text{-N}_2)\text{Zr}(\mu\text{-N}_2)$ produce $[\text{P}_2\text{N}_2]\text{Zr}(\mu\text{-}\eta^2\text{-N}_2\text{H})\text{Zr}(\mu\text{-N}_2)$ ($\mu\text{-H}$) but not the thermodynamically more favorable $[\text{P}_2\text{N}_2]\text{Zr}(\mu\text{-NH})\text{Zr}(\mu\text{-N}_2)$? A theoretical study. *J. Am. Chem. Soc.* **121**, 5754–5761 (1999).
- Basch, H., Musaev, D. G. & Morokuma, K. Can the binuclear dinitrogen complex $[\text{P}_2\text{N}_2]\text{Zr}(\mu\text{-}\eta^2\text{-N}_2)\text{Zr}(\mu\text{-N}_2)$ activate more than one hydrogen molecule? A theoretical study. *Organometallics* **19**, 3393–3403 (2000).
- Janiak, C., Versteeg, U., Lange, K. C. H., Weiman, R. & Hahn, E. The influence of electronic and steric effects and the importance of polymerization conditions in the ethylene polymerization with zirconocene/MAO catalysts. *J. Organomet. Chem.* **501**, 219–234 (1995).
- Sanner, R. D., Manriquez, J. M., Marsh, R. E. & Bercaw, J. E. Structure of μ -dinitrogen-bis(bis(pentamethylcyclopentadienyl)-dinitrogen zirconium(II)), $[(\eta^5\text{-C}_5\text{Me}_5)_2\text{Zr}(\eta^1\text{-N}_2)]_2(\mu_2\text{-N}_2)$. *J. Am. Chem. Soc.* **98**, 8351–8357 (1976).
- Fryzuk, M. D. Activation and functionalization of molecular nitrogen by transition metal complexes. *Chem. Rev.* **3**, 2–11 (2003).
- Manriquez, J. M. et al. Solution structure and dynamics of binuclear dinitrogen complexes of bis(pentamethylcyclopentadienyl) titanium (II) and bis(pentamethylcyclopentadienyl) zirconium (II). *J. Am. Chem. Soc.* **100**, 3078–3083 (1978).
- Manriquez, J. M. & Bercaw, J. E. Preparation of a dinitrogen complex of bis(pentamethylcyclopentadienyl) zirconium (II). Isolation and protonation leading to stoichiometric reduction of dinitrogen to hydrazine. *J. Am. Chem. Soc.* **96**, 6229–6230 (1974).
- Cummins, C. C., Baxter, S. M. & Wolczanski, P. T. Methane and benzene activation via transient $(t\text{-Bu}_3\text{SiNH})_2\text{Zr}=\text{NSi}(t\text{-Bu})_3$. *J. Am. Chem. Soc.* **110**, 8731–8733 (1988).
- Chaney, A. L. & Marbach, E. P. Modified reagents for determination of urea and ammonia. *Clin. Chem.* **8**, 130–132 (1962).
- Hillhouse, G. L. & Bercaw, J. E. Reactions of water and ammonia with bis(pentamethylcyclopentadienyl) complexes of zirconium and hafnium. *J. Am. Chem. Soc.* **106**, 5472–5478 (1984).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the National Science Foundation, the Petroleum Research Fund administered by the American Chemical Society and the National Institutes of Health for funding, Cambridge Isotope Laboratory for the gift of $^{15}\text{N}_2$ and C. A. Bradley for assistance with the indophenolic titrations.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.J.C. (pc92@cornell.edu). X-ray crystallographic coordinates for 1 and 2 have been deposited at the Cambridge Crystallographic Database under numbers 223980 and 223981, respectively.

Low-velocity zone atop the 410-km seismic discontinuity in the northwestern United States

Teh-Ru Alex Song¹, Don. V. Helmberger¹ & Stephen P. Grand²

¹Seismological Laboratory, Division of Geological and Planetary Science, California Institute of Technology, California 91125, USA

²Department of Geological Sciences, UT Austin, Texas 78749, USA

The seismic discontinuity at 410 km depth in the Earth's mantle is generally attributed to the phase transition of $(\text{Mg,Fe})_2\text{SiO}_4$ (refs 1, 2) from the olivine to wadsleyite structure. Variation in the depth of this discontinuity is often taken as a proxy for mantle temperature owing to its response to thermal perturbations. For example, a cold anomaly would elevate the 410-km discontinuity, because of its positive Clapeyron slope, whereas a warm anomaly would depress the discontinuity. But trade-offs between seismic wave-speed heterogeneity and discontinuity topography often inhibit detailed analysis of these discontinuities, and structure often appears very complicated. Here we simultaneously model seismic refracted waves and scattered waves from the 410-km discontinuity in the western United States to constrain structure in the region. We find a low-velocity zone, with a shear-wave velocity drop of 5%, on top of the 410-km discontinuity beneath the northwestern United States, extending from southwestern Oregon to the northern Basin and Range province. This low-velocity zone has a thickness that varies from 20 to 90 km with rapid lateral variations. Its spatial extent coincides with both an anomalous composition of overlying volcanism and seismic 'receiver-function' observations observed above the region. We interpret the low-velocity zone as a compositional anomaly, possibly due to a dense partial-melt layer, which may be linked to prior subduction of the Farallon plate and back-arc extension. The existence of such a layer could be indicative of high water content in the Earth's transition zone.

Spatial variations in topography of the 410-km and the 660-km discontinuities (referred to here as the 410 and 660) are often inferred from SS precursors (length scale of about 1,500–2,000 km)^{3,4}, near subduction zone depth phase precursors⁵, and receiver function analyses (length scale of about 100–300 km)^{6–8}. The latest efforts consider simultaneous inverting for both mantle velocity and discontinuity topography³. In general, the 660 is depressed under cold regions (slabs) as expected, but the 410 appears to be far more complicated^{3–5}. The stacked converted P-to-S phase (Ps) (receiver function) from the 410 is rather weak, complicated and sometimes shows negative pulses above the 410 (ref. 7).

We first compute one-dimensional (1D) full-waveform synthetics⁹, model the direct S-wave triplications at epicentral distances

of 14–17° and explain the timing and amplitude of multiple arrivals coming from fine structures near the 410 (ref. 10). To resolve the trade-offs between discontinuity topography and mantle velocity directly above or below the discontinuity, we model S-wave triplications at epicentral distances of 21–24°. A low-velocity zone (LVZ) on top of the 410 produces a secondary pulse not normally seen at these distances. Using this secondary pulse as a proxy for the existence of an LVZ atop the 410, we examine events located offshore of Washington–Oregon, USA, and recorded by the TriNet broadband network and several temporary PASSCAL broadband arrays (Fig. 1). Perturbed velocity structures are shown at the turning point for a given great-circle path because the sensitivity to 410 structure is greatest there¹¹. The size of the perturbed structure is estimated as the size of the Fresnel zone in both along-path and cross-path directions. We determine a LVZ directly above the 410 in the northwestern US and cross-validate our finding with receiver-function profiles⁷.

Record section A shows typical waveform characteristics sampling this area at epicentral distances of 14–17° (Fig. 2a). It samples the region beneath the California–Oregon border (Fig. 1).

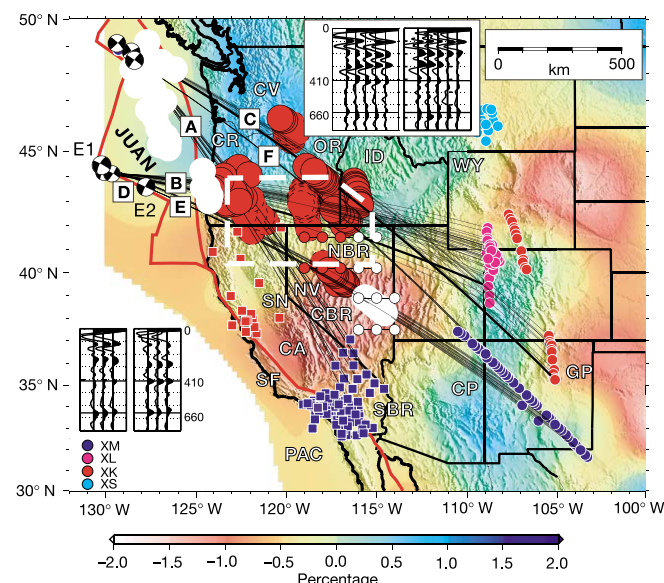


Figure 1 Map of western US displaying samples of fine structure near the 410-km seismic discontinuity. Events off the coast of Oregon and Washington (state borders are in black) are recorded by several PASSCAL arrays (small circles), TriNet (blue squares), and Berkeley network (red squares). Focal mechanisms of these events are shown with black-and-white 'beach balls'. Ray-paths of record sections A–F are shown as black lines. Colour patches display areas sampled by the triplication data; white (normal, no detectable LVZ), large red circles with dashed outline (anomalous 410 with a possible LVZ), large red circles with solid outline (clear LVZ). The areal extent of HAOT13 is indicated by a heavy white dashed line. The base-colour map shows the variation in SV-shear velocity (relative to the 1D reference model) at a depth of 370 km (ref. 12). Thick red lines show the plate boundary. Top inset, receiver-function determination at 41.5°N, 119°–115°W (left) and 40.2°N, 119°–115°W (right). Bottom inset, receiver-function determination at 38.9°N, 116°–114°W (left) and 37.5°N, 116°–114°W (right). Red dots shown in the map (within Nevada) indicate the location where receiver functions show strong negative pulse above the 410, whereas white dots represent the locations where receiver functions show no negative pulse above the 410 and relatively strong simple 410 peaks. GP, Great Plain; WY, Wyoming; NBR, northern Basin and Range; CBR, central Basin and Range; SBR, south Basin and Range; SN, Sierra Nevada; CV, Cascadia Volcanic Arc; CR, Coast Ranges; JUAN, Juan de Fuca plate; PAC, Pacific plate; SF, San Andreas Fault; CA, California; NV, Nevada; OR, Oregon; ID, Idaho; XM, Rio Grande Rift Seismic Transect array (99/07–01/05); XL, Deep Probe array 97/05–97/11; XK, CDROM (Colorado) array (99/04–00/09); XS, Montana Broadband array (99/11–00/09).

Relative to the 1D reference model (Supplementary Information), we observe a considerable increase of separation in timing between the branch AB and the branch CD of about 4–6 s (Fig. 2a, c). At these ranges, the branch AB represents the S wave propagating through the upper mantle and turning at about 200–300 km, and the branch CD represents the S wave bottoming near the 410 (Fig. 2b). Synthetics calculated from a perturbed 1D model with the 410 depressed by 60 km improve the differential time between the AB and CD branches considerably (model Topo in Fig. 2a). However, other models can also fit the data equally well (Fig. 2a, Supplementary Fig. S1).

Record section B (Supplementary Fig. S2) trends more east–west and samples the region beneath the northern Basin and Range including eastern Oregon, northern Nevada and western Idaho (Fig. 1). This record section shows similar waveforms to those in record section A and previously introduced models can be used to fit the data (Fig. 2c, Supplementary Fig. S2). These record sections indicate that the mantle structure near the 410 is relatively slow beneath the northern Basin and Range. The anomalous structure is probably distributed over areas from the California–Oregon border to the northwestern Basin and Range.

Although it is clear that seismic velocity is slow near 410 km depth, we cannot distinguish between these models: (1) a topographic depression of the 410, (2) a modest decrease in velocity gradient above the 410, (3) a LVZ above the 410, or (4) a hybrid model (LVZA) with the 410 depressed and a LVZ atop the 410 (Fig. 2a, Supplementary Fig. 1). It is difficult to explain this anomaly with current three-dimensional (3D) velocity models¹² because variations in shear-wave velocity at these depths are generally small (Fig. 1).

At the longer epicentral distances of 21–24°, however, it is the branch AB that is bottoming near the 410 (Fig. 2b). In combination

with ray-paths at the shorter distances, they provide strong constraints on the structure near the 410. Travel-time curves from model LVZA on top of the 410 not only explain the differential time of branch AB and branch CD at an epicentral distance of 14–17°, but also extend the branch AB from a maximum epicentral distance of 21° to 23° (Fig. 2c).

Record section C (Fig. 3a) samples the northern Basin and Range province and shows an anomalous secondary pulse emerging at epicentral distances of 22–23°, which is also seen in record sections D and F (Fig. 3a, b). The model with a large topographic depression (60 km) of the 410 does not reproduce this feature. The model with a 40-km-thick LVZ containing a 5% reduction in shear velocity produces the observed secondary pulse, but the separation between the first and secondary pulse appears too large (Supplementary Fig. S3). We prefer the hybrid model (LVZA), which explains these record sections in both amplitude and timing. This model has the 410 depressed by 20 km and a 20-km-thick LVZ atop the 410 with a 5% reduction in shear velocity.

Owing to the relatively long period of the S waves (about 6–8 s), a trade-off between the thickness and strength of the LVZ still exists. Therefore, we cannot completely exclude a model having a LVZ of 10 km with 10% velocity drop, or a model having a LVZ of 40 km thick with a 2.5% velocity drop. However, record sections at epicentral distances of 14–17° show that the amplitude of the branch CD predicted by a model having a LVZ with a 10% velocity drop appears too large. Furthermore, receiver-function synthesis (discussed below) indicates that the amplitude of a negative converted pulse due to the existence of a LVZ with only a 2.5% velocity drop cannot explain observed receiver functions. Thus, our preferred model has a 20 km LVZ with a 5% reduction in velocity.

The spatial extent of the LVZ includes regions beneath northern California, Oregon, northern Nevada and southwestern Idaho,

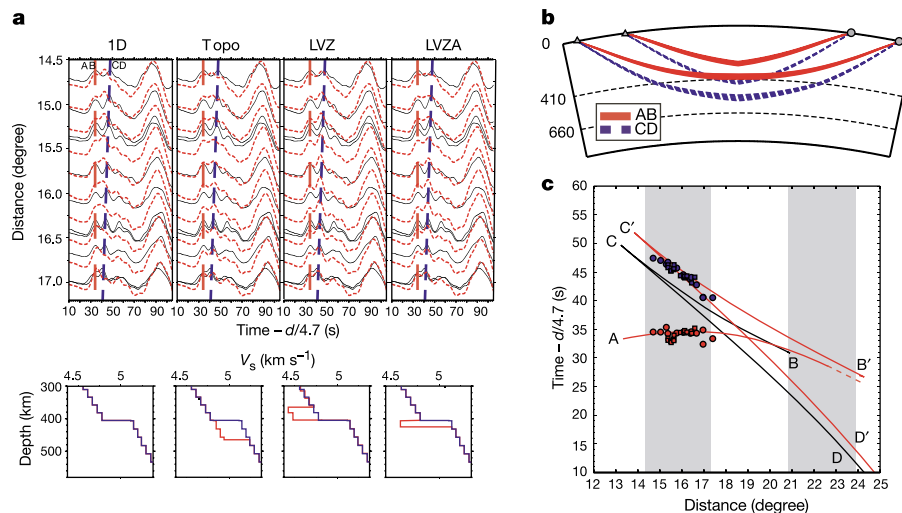


Figure 2 Modelled displacement waveforms (14–17°) with corresponding ray-paths and travel-time triplications. **a**, Record section A (recorded by TriNet, see Fig. 1). Data (solid black traces) are compared with synthetics (dotted red traces) generated from various models, starting from the left: 1D model, model Topo (60 km topography depression), model LVZ (40 km LVZ) and model LVZA (20 km LVZ + 20 km topography depression). Vertical dashed and dash-dotted lines indicate observed triplication branches AB (red) and CD (blue), respectively. Large-amplitude and long-period waves at the end of these traces are Love waves, which are sensitive to the shallow structure. Velocity models used to compute the synthetics are directly shown below these waveforms. The perturbed model is shown as a red line and the reference model is shown as a blue line. The timescale is shown with a reduced velocity of 4.7 km s⁻¹; d is epicentral distance. **b**, Ray-paths of

triplication branches AB (red line) and CD (blue dashed line). The 410 and 660 seismic discontinuities are given as dashed lines. Note that, at the shorter distance (for example, 16°), the branch CD is primarily sensitive to the structure at the turning point near the depth of 400 km. However, at the longer distance (for example, 22°), the branch AB is primarily sensitive to the same structure. **c**, Triplication curves computed from the 1D model (black line) and model LVZA (thick red line). Observed branches AB and CD of record section A (circle) and record section B (square) are projected with red symbols and blue symbols, respectively (see also Fig. 1 for ray-paths). Note that model LVZA delays the branch CD to fit the above observations but also extends the appearance of the branch AB to epicentral distance of 21–23°.

corresponding to an area of about 100,000 km² (Fig. 1). The thickness of the LVZ is about 20 km (Fig. 3a, model LVZA) but appears thicker (70 km thick) beneath the Oregon–Nevada–Idaho border (Fig. 1, Fig. 3a, model LVZB). In central Nevada, the thickness of the LVZ is up to 90 km (Fig. 1, Fig. 3b, model LVZC). However, when examining record section E, we find no evidence for a LVZ atop the 410 beneath the central Basin and Range in eastern Nevada (Fig. 1, Fig. 3b). These observations imply that the variation of the LVZ in thickness is very rapid. This conclusion is supported by receiver function determination⁷, where strongly negative pulses directly above the 410-converted pulses suddenly disappear in profiles east of 116°W (Fig. 1). In addition, we note that record sections sampling regions beneath the Juan de Fuca plate show no signs of a LVZ atop the 410 (Fig. 1).

To validate our result, we produce a synthetic receiver function and compare it to the published result⁷ for areas closest to that sampled by record section F. The synthetic receiver function from the preferred model LVZB indeed produces a negative pulse associated with the LVZ atop the 410, but the predicted onset of the negative pulse is late and the converted pulse from the 410 is too strong (Fig. 3c). The observed receiver-function determination is constructed through stacking multiple receiver-function profiles from various azimuths where both the thickness of the LVZ and the depth of the 410 are probably changing. Under such circumstances, we can produce the complicated converted pulse from the 410 and the negative pulse from the LVZ on top of the 410 (Fig. 3c, see Supplementary Methods). We could not reproduce receiver functions with a strong negative pulse without inserting a LVZ atop the 410.

We note that strongly negative pulses are absent in areas where we detect no LVZ (Fig. 1, large white circles), whereas converted pulses from the 410 are rather strong and simple (Fig. 1). In general, we expect the amplitude of the converted phase from the 410 to be smaller than that from the 660 because of its smaller velocity contrast. One possible explanation is that these relatively strong converted pulses from the 410 could be related to the existence of a

very thin LVZ on top of the 410, which increases the velocity contrast, but if so, it is too thin to be detected.

Geographically, the anomalous region discussed above coincides with a unique type of mantle-derived basalt exposed in the north-western United States, low-potassium, high-alumina olivine theoleiites (HAOT)^{13,14}. Noticeably, HAOT is on average ten times more depleted in incompatible elements than any other type of basalt found in the rest of the Basin and Range and the western USA^{13–16}. It is generally believed that formation and eruption of HAOT are strongly linked to extensional tectonics in the back-arc regime. The geographical coincidence between the unique HAOT and the LVZ atop the 410 in the northwestern Basin and Range might suggest that they are probably linked to tectonic processes such as subduction of the Farallon plate and surface extension in the back-arc regime¹⁷. Other reports of such LVZs directly above the 410 are usually associated with back-arc regimes of ongoing subduction or ancient subduction^{5,18–21}.

If the LVZ is primarily due to variations in mantle temperature, the thickness of the LVZ is expected to vary gradually. However, the thickness of the LVZ varies rapidly over length scales of about 100–200 km. Therefore, we argue against a thermally dominant origin and prefer a compositional anomaly. The presence of partial melt could considerably lower the velocity²² but this interpretation is directly linked to the water content actually brought into the deep mantle^{23,24}. If the water content is sufficiently high in the transition zone, the LVZ may form directly above the 410 in upwelling material as a result of dehydration melting²⁵.

For an anhydrous silicate melt, the density of melt can be equal to or larger than that of ambient mantle at depths close to the 410 (refs 26–29). For a hydrous silicate melt, however, not much information is currently available on its density. If the effect of water on the bulk modulus of solids is considered similar to that on the bulk modulus of melts, while the linear relationship between bulk modulus and density holds³⁰, then hydrous melts may be even denser than hydrous solids. Therefore, the LVZs would be relatively stable for long periods of time. □

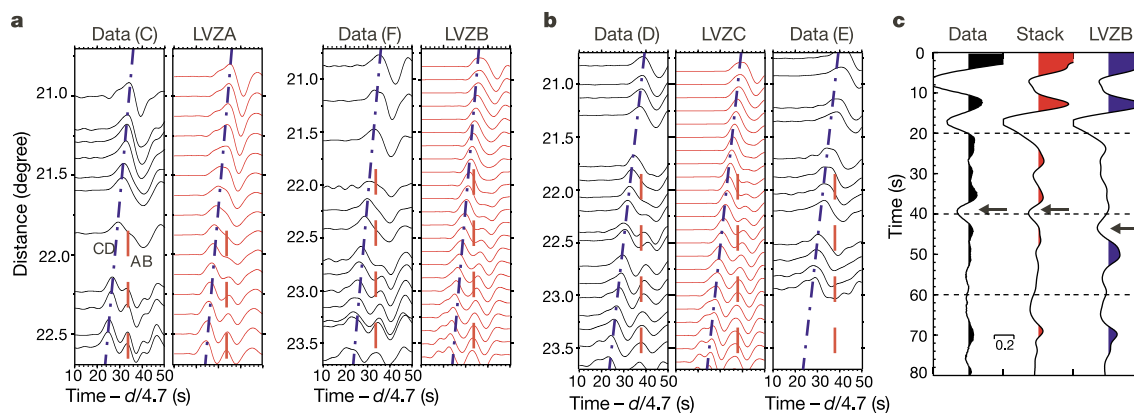


Figure 3 Modelled velocity waveforms (21–24°) and receiver function profiles. **a**, Record section C (recorded by XK PASSCAL stations; see Fig. 1) and synthetics from a hybrid model LVZA (20 km LVZ + 20 km topography depression) are shown in the left two columns. Record section F (recorded by XM PASSCAL stations; see Fig. 1) and synthetics from a hybrid model LVZB (70 km LVZ + 60 km topography depression) are shown in the right two columns. Note that the separation between branches AB and CD is larger in record section F than that in record section C, indicating a thicker LVZ. **b**, Record section D (recorded by XM PASSCAL stations; see Fig. 1) and synthetics from a hybrid model LVZC (90 km LVZ + 60 km topography depression) are shown in the left two columns. Record section E (recorded by XM PASSCAL stations) is shown in the right column. In **a** and **b**, vertical dashed and dash-dotted lines indicate observed triplication branches AB (red) and CD (blue), respectively (see also Fig. 2b). The timescale is shown with a reduced velocity of

4.7 km s⁻¹; *d* is epicentral distance. Triplication branch bottoming beyond the 660 arrives at or even ahead of the CD branch at these ranges. Note record sections D and E sample regions only 100 km apart (see also Fig. 1). No secondary pulse branch AB is observed in record section E at these distances. **c**, Comparison of stacking receiver-function profile⁷ and forward calculations. All traces are normalized by the amplitude of the P wave and shown with a ray parameter of 0.06 s km⁻¹. The left trace shows the receiver-function profile, which samples similar regions to record section F (see also Fig. 1). The right trace shows the synthetic receiver function from model LVZB. The middle trace shows a synthetic receiver function by stacking receiver functions computed from multiple 1D models, which have variations in the 410 and the thickness of LVZ (see Supplementary Methods). The arrow points out the negative converted pulse caused by the LVZ on top of the 410.

Received 12 August; accepted 14 November 2003; doi:10.1038/nature02231.

1. Helffrich, G. Topography of the transition zone seismic discontinuity. *Rev. Geophys.* **38**, 141–158 (2000).
2. Shearer, P. M. et al. in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale* (ed. Karato, S.-I.) 115 (AGU Geophys. Monogr. Vol. 117, Washington DC, 2000).
3. Gu, Y., Dziwonski, A. M. & Ekstrom, G. Simultaneous inversion for mantle velocity and topography of transition zone discontinuities. *Geophys. J. Int.* **154**, 559–583 (2003).
4. Flanagan, M. P. & Shearer, P. M. Global mapping of topography on transition zone velocity discontinuities by stacking SS precursors. *J. Geophys. Res.* **103**, 2673–2692 (1998).
5. Flanagan, M. P. & Shearer, P. M. Topography on the 410 km seismic velocity discontinuity near subduction zones from stacking of sS, sP, and pP precursors. *J. Geophys. Res.* **103**, 21165–21182 (1998).
6. Ramesh, D. S., Kind, R. & Yuan, X. Receiver function analysis of the North America crust and upper mantle. *Geophys. J. Int.* **150**, 91–108 (2002).
7. Gilbert, H. J., Sheehan, A. F., Dueker, K. G. & Molnar, P. Receiver functions in the western United States, with implications for upper mantle structure and dynamics. *J. Geophys. Res.* **109**(B5), doi:10.1029/2001JB001194 (2003).
8. Dueker, K. G. & Sheehan, A. F. Mantle discontinuity structure from midpoint stacks of converted P to S waves across the Yellowstone hotspot track. *J. Geophys. Res.* **102**, 8313–8327 (1997).
9. Saikia, C. K. Modified frequency-wave-number algorithm for regional seismograms using Filons quadrature-modeling of Lg waves in eastern North America. *Geophys. J. Int.* **118**, 142–158 (1994).
10. Grand, S. P. & Helmberger, D. V. Upper mantle shear structure of North America. *Geophys. J. R. Astron. Soc.* **76**, 399–438 (1984).
11. Kennett, B. L. N. & Bowman, J. R. The velocity structure and heterogeneity of the upper mantle. *Phys. Earth Planet. Inter.* **59**, 134–144 (1990).
12. Van der Lee, S. High-resolution estimates of lithospheric thickness from Missouri to Massachusetts, USA. *Earth Planet. Sci. Lett.* **203**, 15–23 (2002).
13. Hart, W. K., Aronson, J. L. & Mertzman, S. A. Areal distribution and age of low-K, high-alumina olivine tholeiite magmatism in the northwestern Great Basin. *Geol. Soc. Am. Bull.* **95**, 186–195 (1984).
14. Hart, W. K. Chemical and isotopic evidence for mixing between depleted and enriched mantle, northwestern USA. *Geochim. Cosmochim. Acta* **49**, 131–144 (1985).
15. Fitton, J. G., James, D. & Leeman, W. P. Basic magmatism associated with late Cenozoic extension in the western United States: Compositional variations in space and time. *J. Geophys. Res.* **96**, 13693–13711 (1991).
16. Kempton, P. D., Fitton, J. G., Hawkesworth, C. J. & Ormerod, D. S. Isotopic and trace element constraints on the composition and evolution of the lithosphere beneath the southwestern United States. *J. Geophys. Res.* **96**, 13713–13735 (1991).
17. Awater, T. & Stock, J. Pacific North America plate tectonics of the Neogene southwestern United States. *Int. Geol. Rev.* **40**, 375–402 (1998).
18. Sacks, I. S. & Snoke, J. A. The use of converted phases to infer the depth of the lithosphere-asthenosphere boundary beneath South America. *J. Geophys. Res.* **82**, 2011–2017 (1977).
19. Revenaugh, J. & Sipkin, S. A. Seismic evidence for silicate melt atop the 410-km mantle discontinuity. *Nature* **369**, 474–476 (1994).
20. Nolet, G. & Zielhuis, A. Low S velocities under the Tornquist-Teisseyre zone: evidence for water injection into the transition zone by subduction. *J. Geophys. Res.* **99**, 15813–15820 (1994).
21. Vinnik, L., Kumar, M. R., Kind, R. & Farra, V. Super-deep low-velocity layer beneath the Arabian plate. *Geophys. Res. Lett.* **30**(1415), doi:10.1029/2002GL016590 (2003).
22. Hammond, W. C. & Humphreys, E. D. Upper mantle seismic wave velocity: Effects of realistic partial melt geometries. *J. Geophys. Res.* **105**, 10975–10986 (2000).
23. Williams, Q. & Hemley, R. J. Hydrogen in the deep Earth. *Annu. Rev. Earth Planet. Sci.* **29**, 365–418 (2001).
24. Dixon, J. E., Leist, L., Langmuir, J. & Schilling, J.-G. Recycled dehydrated lithosphere observed in plume-influenced mid-ocean-ridge basalt. *Nature* **420**, 385–389 (2002).
25. Bercovici, D. & Karato, S. Whole-mantle convection and the transition-zone water-filter. *Nature* **425**, 39–44 (2003).
26. Stolper, E. M., Walker, D., Hager, B. H. & Hays, J. F. Melt segregation from partially molten source regions: the importance of melt density and source region size. *J. Geophys. Res.* **86**, 6261–6271 (1981).
27. Rigden, S. M., Ahrens, T. J. & Stolper, E. M. Densities of liquid silicates at high pressure. *Science* **226**, 1071–1074 (1984).
28. Agee, C. B. Crystal-liquid density inversions in terrestrial and lunar magmas. *Phys. Earth Planet. Inter.* **107**, 63–74 (1998).
29. Chen, G. Q., Ahrens, T. J. & Stolper, E. M. Shock-wave equation of state of molten and solid fayalite. *Phys. Earth Planet. Inter.* **134**, 35–52 (2002).
30. Angel, R. J., Frost, D. J., Ross, N. L. & Hemley, R. Stabilities and equations of state of dense hydrous magnesium silicates. *Phys. Earth Planet. Inter.* **127**, 181–196 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Ni, R. Aster, and the rest of the RISTRA team for the use of their data set, and we acknowledge all who have contributed to the PASSCAL arrays and IRIS DMC. We also thank H. Gilbert for providing receiver-function profiles. Reviews from D. Anderson, P. Asimow, T. Ahrens, B. Savage and M. Simons also helped to clarify early versions of this paper. We benefited from discussions with H. Gilbert, D. Anderson, H. Kanamori, Y. Tan, S. Ni and M. Gurnis. We thank L. Zhu for sharing his receiver function modelling code. This study was supported by the NSF and is a contribution to Division of Geological and Planetary Science, Caltech.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.-R.A.S. (alex@gps.caltech.edu).

Mangroves enhance the biomass of coral reef fish communities in the Caribbean

Peter J. Mumby¹, Alasdair J. Edwards², J. Ernesto Arias-González³, Kenyon C. Lindeman⁴, Paul G. Blackwell⁵, Angela Gall², Malgosia I. Górczyska², Alastair R. Harborne¹, Claire L. Pescod², Henk Renken¹, Colette C. Wabnitz⁶ & Ghislane Llewellyn⁷

¹Marine Spatial Ecology Laboratory, School of Biological and Chemical Sciences, University of Exeter, Prince of Wales Road, Exeter EX4 4PS, UK

²School of Biology, Ridley Building, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK

³Laboratorio Ecología de Ecosistemas de Arrecifes Coralinos, Departamento Recursos del Mar, CINVESTAV Unidad Mérida, AP73 Cordemex, CP97310 Mérida, Yucatán, Mexico

⁴Environmental Defense, Caribbean Field Office, Miami, Florida 33186, USA

⁵Department of Probability and Statistics, University of Sheffield, Hicks Building, Sheffield S3 7RH, UK

⁶Fisheries Centre, Lower Mall Research Station, The University of British Columbia, Vancouver, British Columbia V6T1Z4, Canada

⁷Conservation Science Program, WWF-US, 1250 24th Street Northwest, Washington DC 20037, USA

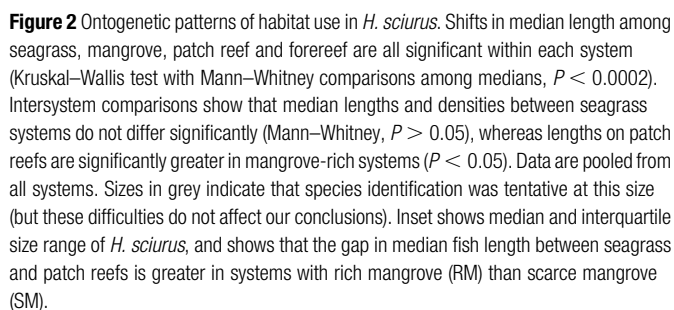
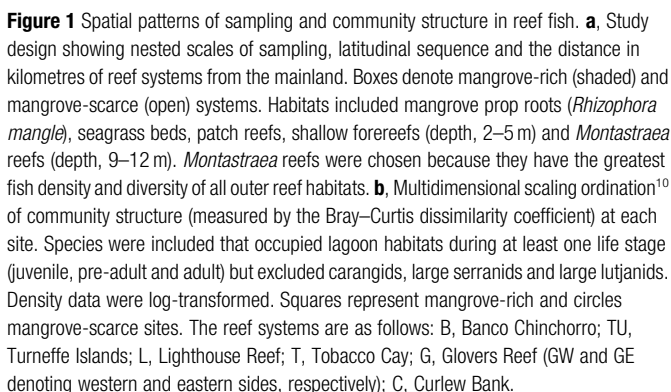
Mangrove forests are one of the world's most threatened tropical ecosystems with global loss exceeding 35% (ref. 1). Juvenile coral reef fish often inhabit mangroves^{2–5}, but the importance of these nurseries to reef fish population dynamics has not been quantified. Indeed, mangroves might be expected to have negligible influence on reef fish communities: juvenile fish can inhabit alternative habitats and fish populations may be regulated by other limiting factors such as larval supply or fishing⁶. Here we show that mangroves are unexpectedly important, serving as an intermediate nursery habitat that may increase the survivorship of young fish. Mangroves in the Caribbean strongly influence the community structure of fish on neighbouring coral reefs. In addition, the biomass of several commercially important species is more than doubled when adult habitat is connected to mangroves. The largest herbivorous fish in the Atlantic, *Scarus guacamaia*, has a functional dependency on mangroves and has suffered local extinction after mangrove removal. Current rates of mangrove deforestation are likely to have severe deleterious consequences for the ecosystem function, fisheries productivity and resilience of reefs. Conservation efforts should protect connected corridors of mangroves, seagrass beds and coral reefs.

The Mesoamerican reef system of Belize and Mexico provides a unique experimental setting that has allowed us to isolate the importance of mangroves to coral reef fish. Three atolls have virtually no, or extremely limited, mangrove cover. As migrations from the nearest mangrove resource, across 10–25 km of open ocean with depths exceeding 2,000 m, are likely to be insignificant for demersal reef species⁷, we can assume that adult fish must have used nursery habitats, such as seagrass, on the atolls. These reef systems provide three 'scarce mangrove' treatments, in which the mean mangrove perimeter is only 3.9 km within an area the size of Glovers Reef (228 km²).

Uniquely in the region, Belize also possesses a mangrove-dominated atoll and extensive offshore mangrove islands at the edge of a barrier reef. The offshore barrier reef is separated from the mainland by a channel that is roughly 15-km wide, and sediment cores show that there is little (<1%) connectivity with the mainland⁸. The existence of these offshore, 'rich mangrove' atoll and barrier reef areas allowed us to contrast the fish communities of three mangrove-scarce reef systems with those of three mangrove-rich systems (one atoll and two areas of the barrier reef). The

The structure of reef fish communities in outer *Montastraea* reefs differed markedly between mangrove-rich and mangrove-scarce sites (Fig. 1b). The magnitude of such differences was tested by nested ANOSIM¹⁰, where the output statistic (*R*) equals unity when a factor, such as mangrove extent, divides the data into tight, non-overlapping groups. Given the possibility that mangroves are a redundant nursery habitat and only one of a range of potential limiting factors, we were surprised to find that mangrove extent was a dominant factor structuring reef fish communities. Mangrove

Studies elsewhere in the Caribbean suggest that four reef fish species are heavily dependent on lagoonal nurseries, although, unlike in our study, the importance of mangroves was not isolated¹¹. These species were the striped parrotfish (*Scarus iserti*), bluestriped grunt (*Haemulon sciurus*) and the commercially important yellowtail (*Ocyurus chrysurus*) and schoolmaster (*Lutjanus apodus*) snap-



pers. We therefore examined the biomass of these species, but we included two additional species that were often seen in the mangroves as juveniles: the French grunt (*Haemulon flavolineatum*) and white grunt (*Haemulon plumieri*). Although none of the species was absent from reefs in mangrove-scarce systems, their biomass was significantly enhanced in at least one reef habitat in mangrove-rich systems (Table 1). The magnitude and pattern of biomass enrichment differed among species (Table 1). *H. sciurus* benefited most strongly from mangroves: biomass on patch reefs in mangrove-rich systems was over 25 times higher than that in mangrove-scarce systems. The biomass of *O. chrysurus* doubled when its preferred *Montastraea* habitat was adjacent to rich mangroves (Table 1). A similar analysis for *S. iserti* showed a 42% biomass increase on the *Montastraea* forereef. *L. apodus* and *H. flavolineatum* biomasses were significantly enriched in patch reef and shallow forereefs but not in the outer *Montastraea* reef.

Mangroves may enhance adult fish biomass in two ways. First, efflux of detritus and nutrients may enrich primary production in neighbouring ecosystems; however, this hypothesis is not well supported¹². Second, mangrove nurseries may provide a refuge from predators and/or plentiful food that increases the survivorship of juveniles¹³. Our data, although not constituting proof, support the latter hypothesis. For example, the size–frequency distribution of *H. sciurus* suggests an ontogenetic shift in habitat use from seagrass, to mangroves, to patch reefs, and finally to forereefs, their main adult habitat (Fig. 2). Our data suggest that juvenile grunts migrate from seagrass beds when they reach a length of 4–6 cm.

Migration occurs from seagrass to mangroves, but if mangroves are absent the grunts move to reefs (Fig. 2). Because mangroves offer refuge¹³ and the biomass of haemulid predators is greater on reefs than in mangroves (30 t km⁻² versus 18 t km⁻²; *t*-test, *P* < 0.05), the chances of grunt survival may be lower if grunts migrate directly to reefs. In short, some fish species move to their adult habitat in stages. As the biomass of predators increases at each stage, it is desirable to grow as large as possible before taking the next step towards adult habitat. We suggest that mangroves provide an intermediate nursery stage between seagrass beds and patch reefs, and they therefore alleviate a predatory bottleneck in early demersal ontogeny. For further details see Supplementary Information.

Large-scale ecological studies of pattern should infer causation only if alternative causative mechanisms have been tested and falsified¹⁴. Thus, we can only infer that mangroves enrich reef fish communities once plausible alternative explanations, such as variations in fishing intensity, have been discounted. There seemed to be four plausible alternative scenarios, each of which was tested and discounted (Table 2). Indeed, the pattern of direct fishing pressure tended to strengthen our conclusions: lutjanid biomasses were greater in mangrove-rich areas despite higher fishing pressure (Table 2). Therefore, either fishing of lutjanids is not particularly intense or we have underestimated the potential biomass enrichment of lutjanids by mangroves.

None of the 44 fish species that we encountered in mangroves was confined to that habitat and 37 (84%) were seen on *Montastraea* reefs as reproductively capable adults. However, juveniles of one reef

Table 1 Impact of extensive mangroves on the biomass of fish in patch reef, shallow forereef and *Montastraea* reef habitats

Species	Mean (s.e.m.) biomass (kg km ⁻²)											
	Patch reef				Shallow forereef				<i>Montastraea</i> reef			
	Scarce mangroves	Rich mangroves	Sig. factors*§	Biomass increase	Scarce mangroves	Rich mangroves	Sig. factors*	Biomass increase	Scarce mangroves	Rich mangroves	Sig. factors*	Biomass increase
<i>S. iserti</i>	—	—	—	—	—	—	—	—	1,530 (118)	2,170 (111)	M‡	42%
<i>H. sciurus</i>	1,205 (329)	33,349 (9,274)	M	2667%	56 (38)	425 (120)	M	659%	288 (53)	447 (55)	M, R	55%
<i>H. flavolineatum</i>	5,256 (1,460)	15,307 (4,114)	M, R	191%	516 (101)	1,600 (249)	M	210%	1,398 (149)	1,643 (139)	NS	—
<i>H. plumieri</i>	5,174 (1,614)	16,280 (3,591)	M	214%	317 (72)	843 (304)	M	165%	523 (62)	863 (69)	M	65%
Haemulidae¶	11,636 (2,089)	67,370 (12,971)	M, R	478%	889 (152)	3,031 (497)	M	240%	2,288 (188)	3,210 (192)	M, R, S‡	40%
<i>O. chrysurus</i>	769 (441)	410 (95)	R†	—	659 (150)	892 (187)	NS	—	3,098 (486)	6,715 (1,323)	M, S‡	116%
<i>L. apodus</i>	739 (354)	6,192 (1,566)	M	737%	622 (336)	2,392 (722)	M	284%	1,767 (226)	1,898 (259)	S‡	—
Lutjanidae¶	2,890 (1,228)	16,707 (4,805)	M	478%	1,882 (745)	4,428 (1,055)	M	135%	5,883 (796)	12,223 (1,503)	M, S‡	107%

The importance of mangroves and reef system in explaining patterns of biomass was tested by nested general linear model ANOVA. Data were transformed by the Box–Cox method. NS, not significant.

*Significant factors (*P* < 0.05) are mangrove (M), reef system (R) and site (S). Where mangroves exerted significant influence, the increase in mean biomass is expressed as a percentage of the level in mangrove-scarce systems. Neither the biomass nor the density of any species in seagrass beds differed significantly between mangrove-rich and mangrove-scarce systems.

†The fit by the Box–Cox method was poor (*P* > 0.02), and the significance of mangroves could not be tested.

‡Site was only entered into these tests because site-level data (individual transects) often had to be pooled to increase sample size.

§Patch reef area was entered as a covariate for patch reefs but the slopes did not differ from zero.

|| *S. iserti* was not surveyed on patch reefs or shallow forereefs.

¶ All species in the family were assessed, not just those shown in the table.

Table 2 Alternative explanations of increased biomass observed in some haemulids, scarids and lutjanids in *Montastraea* reefs adjacent to rich mangroves

Potential alternative explanation of results	Test	Result	Decision
Direct fishing pressure: greater fishing pressure on haemulids and lutjanids in <i>M_{scarce}</i> systems	Compare density of artisanal and commercial fishers between <i>M_{rich}</i> and <i>M_{scarce}</i> systems*	Greater in all <i>M_{rich}</i> systems (J. Azueta, personal communication)	Reject
Indirect fishing pressure: fishing led to fewer predators of haemulids and lutjanids in <i>M_{rich}</i> systems	ANOVA of piscivore biomass between <i>M_{rich}</i> and <i>M_{scarce}</i> systems	No difference (10.5 ± 2.9 tonnes km ⁻² versus 10.1 ± 1.6 t km ⁻² ; mean ± s.e.m.)	Reject
Subtleties in habitat type: structure of reef habitat differs between <i>M_{rich}</i> and <i>M_{scarce}</i> systems	Nested ANOSIM of habitat composition and rugosity variables <i>M_{rich}</i> versus <i>M_{scarce}</i>	No difference (<i>P</i> > 0.05)	Reject
Internal validity of design: combination of barrier reef and atoll systems for <i>M_{rich}</i> treatment caused bias	Compare barrier and atoll systems for community structure and focal species biomass (evidence of bias), benthic structure (δ habitat), density of early haemulids in seagrass (δ recruitment)	Data do not group into barrier and atoll categories. All nested ANOVAs and ANOSIMs not significant (<i>P</i> > 0.05). Mean haemulid densities 78 ha ⁻¹ and 79 ha ⁻¹ in atoll and barrier seagrass	Reject

Most of these analyses focus on the *Montastraea* habitat, which, being most distant from mangroves, is perhaps most likely to be confounded by other factors. M represents the factor Mangrove entered into the analysis. δ denotes 'difference'.

*Fisheries statistics are currently being collated by J. Azueta (Fisheries Department, Belize), but the differences between *M_{rich}* and *M_{scarce}* reef systems are unequivocal and relate to the distribution of inhabited islands, fishing camps and proximity to fish markets.

species, the rainbow parrotfish *S. guacamaia*, were found exclusively in mangroves. *S. guacamaia* is the largest herbivorous marine fish in the Atlantic¹⁵ (reaching 1.2 m in length) and is listed as vulnerable on the IUCN *Red List of Threatened Species*¹⁶. Rates of encounter of *S. guacamaia* juveniles were very significantly greater in mangroves than in any other habitat ($P < 0.0001$), and we conclude that they are dependent on mangroves. This is consistent with limited reports indicating that *S. guacamaia* juveniles are usually seen in mangroves^{15,17}. If juveniles are mangrove dependent, we would expect adult *S. guacamaia* to be scarce or absent on reefs with little access to mangroves. This was found to be true: adult densities of *S. guacamaia* were strongly and significantly ($P < 0.0001$) enhanced on reefs near mangroves (see Supplementary Information).

Such functional dependency means that *S. guacamaia* is vulnerable to local extinction from habitat loss as well as from overfishing. Indeed, anecdotal information from Glovers Reef (D. Wesby, personal communication) suggests that *S. guacamaia* has undergone local extinction in the past 30 yr. Schools of this parrotfish were commonly observed in the 1960s when several of the islands had well-developed mangrove habitats. Unlike other study sites, all functional mangrove was cleared in the late 1960s and early 1970s, and in the mid- to late 1970s *S. guacamaia* was heavily fished. *S. guacamaia* is no longer fished and either has recovered or has survived at low densities at mangrove-rich sites. Its extinction at Glovers Reef seems most probably due to the removal of its nursery habitat.

The impact of historical overfishing on modern reef ecosystems has been discussed at length¹⁸. Reductions in herbivory may reduce the resilience¹⁹ of coral reefs to algal overgrowth. In the case of *S. guacamaia*, historical overfishing and mangrove deforestation may have worked synergistically to reduce herbivory and secondary production on many Caribbean coral reefs. We estimate that loss of a single adult *S. guacamaia* would constitute a 10% reduction in total parrotfish biomass within its territory (see Supplementary Information).

Extensive mangrove habitats can enhance the biomass of fishes on Caribbean reefs because tropical coastal ecosystems are functionally linked. Although precise corridors of connectivity between habitats are not fully understood as yet, the results have an important implication for conservation planning: management schemes should explicitly protect swaths of connected habitats rather than simply identify representative areas of each habitat in isolation²⁰. Given the ever-increasing range and severity of natural and anthropogenic disturbances to coral reefs²¹, any natural source of ecosystem production and resilience should be conserved. Our data suggest that the current rate of mangrove deforestation, which is greatest in the Americas at $2,251 \text{ km}^2 \text{ yr}^{-1}$ and exceeds that of tropical rainforests¹, will have significant deleterious consequences for the functioning, fisheries, biodiversity and resilience of Caribbean coral reefs. □

Methods

Fish and benthic surveys

Montastraea reefs were surveyed at a depth of 9–12 m, just inside the reef escarpment. All but nocturnal (such as Apogonidae) and highly cryptic (such as Clinidae and Gobiidae) fish species were surveyed by using discrete group visual fish census²². Those species were also ignored in other comparable studies²³. Species were divided into four groups, and their density and size (to the nearest centimetre) were estimated along belt transects by the same person at each site. Surveys were carried out at 09:00, 12:00 and 15:00 without systematic bias per site.

Transect size and number were optimized by using species-area curves from pilot surveys at Glovers Reef, which had relatively low fish density. The transect dimensions and numbers (given in parentheses) at each site were $30 \times 2 \text{ m}$ (6) for smaller benthic species; $30 \times 4 \text{ m}$ (10) for scarids, acanthurids, pomacanthids, diodontids and monacanthids; $100 \times 4 \text{ m}$ (6) for haemulids, chaetodontids, small serranids and labrids; and $100 \times 6 \text{ m}$ (6) for lutjanids, carangids, planktivorous labrids, large serranids and other large predators. Lutjanids and haemulids were surveyed on the shallow fore reef by using $150 \times 4 \text{ m}$ (6) transects and on patch reefs by using $10 \times 4 \text{ m}$ (4) transects. We surveyed seagrass beds by using $50 \times 2 \text{ m}$ (12) transects and mangrove fringes by using $20 \times 2 \text{ m}$ (10) transects, of which less than a metre extended outside prop roots.

Fish lengths were converted to biomass by using allometric relationships²⁴. The percentage cover of coral, macroalgae, turf algae, coralline red algae and sand was measured in fifteen 0.25-m^2 quadrats per site. As an index of structural complexity, we measured the horizontal distance covered by a 2-m chain (0.3-cm links) draped tightly across the seabed ($n = 10$ per site). All surveys were completed during May and June 2002.

Calculation of predator biomass and statistical analysis

Previous studies have found that the main predators of haemulids and smaller lutjanids are *Sphyrna barracuda*, large lutjanids (such as *L. jocu*) and serranids (such as *Mycteroperca bonaci*) and carangids (such as *Caranx hippos*). We pooled biomass data for these groups to obtain a proxy for predation, which we tested by nested general linear model analysis of variance (ANOVA; site nested within reef, reef nested within mangrove extent).

Observations of uncommon species such as *S. guacamaia* cannot be analysed by conventional statistical methods, so we used bayesian methods based on the number of observations rather than fish densities (see Supplementary Information). Data were pooled from all six reef systems surveyed.

Received 8 August; accepted 9 December 2003; doi:10.1038/nature02286.

- Valiela, I., Bowen, J. L. & York, J. K. Mangrove forests: One of the world's threatened major tropical environments. *Bioscience* **51**, 807–815 (2001).
- Rooper, J. R. & Dennis, G. D. Diel, lunar and seasonal changes in a mangrove fish assemblage off southwestern Puerto Rico. *Bull. Mar. Sci.* **49**, 684–698 (1991).
- Claro, R. & García-Arteaga, J. P. Estructura de las comunidades de peces asociados a los manglares del grupo insular sabana-camaguey. *Cuba. Avicennia* **10**, 60–83 (1993).
- Ley, J. A., McIvor, C. C. & Montague, C. L. Fishes in mangrove prop-root habitats of northeastern Florida bay: distinct assemblages across an estuarine gradient. *Estuar. Coast. Shelf Sci.* **48**, 701–723 (1999).
- Nagelkerken, I. *et al.* Dependence of Caribbean reef fishes on mangroves and seagrass beds as nursery habitats: a comparison of fish faunas between bays with and without mangroves/seagrass beds. *Mar. Ecol. Prog. Ser.* **214**, 225–235 (2001).
- Caley, M. J. *et al.* Recruitment and the local dynamics of open marine populations. *Annu. Rev. Ecol. System.* **27**, 477–500 (1996).
- Chapman, M. R. & Kramer, D. L. Movements of fishes within and among fringing coral reefs in Barbados. *Environ. Biol. Fish.* **57**, 11–24 (2000).
- Purdy, E. G. & Gischler, E. The Belize margin revisited: 1. Holocene marine facies. *Int. J. Earth Sci.* **92**, 532–551 (2003).
- Sullivan-Sealey, K. & Bustamante, G. *Setting Geographic Priorities for Marine Conservation in Latin America and the Caribbean* (Arlington, VA, The Nature Conservancy, 1999).
- Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**, 117–143 (1993).
- Nagelkerken, I. *et al.* How important are mangroves and seagrass beds for coral-reef fish? The nursery hypothesis tested on an island scale. *Mar. Ecol. Prog. Ser.* **244**, 299–305 (2002).
- Ogden, J. C. in *Life and Death of Coral Reefs* (ed. Birkeland, C.) 288–297 (Chapman and Hall, New York, 1997).
- Laegdsgaard, P. & Johnson, C. Why do juvenile fish utilise mangrove habitats? *J. Exp. Mar. Biol. Ecol.* **257**, 229–253 (2001).
- Manly, B. F. J. *The Design and Analysis of Research Studies* (Cambridge Univ. Press, Cambridge, 1992).
- Randall, J. E. Food habits of reef fishes of the West Indies. *Stud. Trop. Oceanogr.* **5**, 665–847 (1967).
- International Union for Conservation of Nature and Natural Resources (IUCN) 2002 *IUCN Red List of Threatened Species* (<http://www.redlist.org/>) (2002).
- Seraphy, J. E., Faunce, C. H. & Lorenz, J. J. *Mangrove shoreline fishes of Biscayne bay* (Bull. Mar. Sci., Florida).
- Pandolfi, J. M. *et al.* Global trajectories of the long-term decline of coral reef ecosystems. *Science* **301**, 955–958 (2003).
- Gunderson, L. H. Ecological resilience—in theory and application. *Annu. Rev. Ecol. System.* **31**, 425–439 (2000).
- McNeill, S. E. The selection and design of marine protected areas: Australia as a case study. *Biodiv. Conserv.* **3**, 586–605 (1994).
- Knowlton, N. The future of coral reefs. *Proc. Natl Acad. Sci. USA* **98**, 5419–5425 (2001).
- Green, L. E. & Alevizon, W. S. Comparative accuracies of visual assessment methods for coral reef fishes. *Bull. Mar. Sci.* **44**, 899–912 (1989).
- Nagelkerken, I., Dorenbosch, M., Verberk, W., de la Morinière, E. C. & van der Velde, G. Importance of shallow-water biotopes of a Caribbean bay for juvenile coral reef fishes: Patterns in biotope association, community structure and spatial distribution. *Mar. Ecol. Prog. Ser.* **202**, 175–192 (2000).
- Bohnsack, J. A. & Harper, D. E. *Length-Weight Relationships of Selected Marine Reef Fishes from the Southeastern United States and the Caribbean Technical Memorandum NMFS-SEFC-215* (National Oceanic and Atmospheric Administration, Miami, Florida, 1988).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Green for help with the online figures; B. Cowen, J. Pandolfi, S. Palumbi, D. Snyder, F. Micheli, B. Brown and J. Bythell for comments on the manuscript. We thank the National Geographic Society, the US World Wildlife Fund (WWF-US) and The Royal Society for funding. Field work was supported in part by the Lighthouse Reef Resort.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.J.M. (p.j.mumby@exeter.ac.uk).

Mutations in *VKORC1* cause warfarin resistance and multiple coagulation factor deficiency type 2

Simone Rost^{1,2*}, Andreas Fregin^{1*}, Vytautas Ivaskevicius³, Ernst Conzelmann⁴, Konstanze Hörtnagel², Hans-Joachim Pelz⁵, Knut Lappégard⁶, Erhard Seifried³, Inge Scharrer⁷, Edward G. D. Tuddenham⁸, Clemens R. Müller¹, Tim M. Strom^{2,9} & Johannes Oldenburg^{1,3}

¹Department of Human Genetics, University of Würzburg, Biozentrum, Am Hubland, 97074 Würzburg, Germany

²Institute of Human Genetics, GSF National Research Center, Ingolstädter Landstrasse 1, 85764 München-Neuherberg, Germany

³Institute of Transfusion Medicine and Immune Haematology of the DRK Blood Donor Service, Sandhofstrasse 1, Johann Wolfgang Goethe-Universität, 60526 Frankfurt, Germany

⁴Department of Physiological Chemistry II, University of Würzburg, Biozentrum, Am Hubland, 97074 Würzburg, Germany

⁵Federal Biological Research Center for Agriculture and Forestry, Institute for Nematology and Vertebrate Research, Toppeideweg 88, 48161 Münster, Germany

⁶Department of Medicine, Nordland Hospital, 8092 Bodo, Norway

⁷Center of Internal Medicine, Johann Wolfgang Goethe-Universität, Theodor-Stern-Kai 7, 60528 Frankfurt, Germany

⁸MRC Clinical Sciences Centre, Imperial College, London W12 0NN, UK

⁹Institute of Human Genetics, Klinikum rechts der Isar, Technical University, 81675 München, Germany

* These authors contributed equally to this work

Coumarin derivatives such as warfarin represent the therapy of choice for the long-term treatment and prevention of thromboembolic events. Coumarins target blood coagulation by inhibiting the vitamin K epoxide reductase multiprotein complex (VKOR)¹. This complex recycles vitamin K 2,3-epoxide to vitamin K hydroquinone, a cofactor that is essential for the post-translational γ -carboxylation of several blood coagulation factors^{2,3}. Despite extensive efforts, the components of the VKOR complex have not been identified^{4–8}. The complex has been proposed to be involved in two heritable human diseases: combined deficiency of vitamin-K-dependent clotting factors type 2 (VKCFD2; Online Mendelian Inheritance in Man (OMIM) 607473), and resistance to coumarin-type anticoagulant drugs (warfarin resistance, WR; OMIM 122700). Here we identify, by using linkage information from three species, the gene *vitamin K epoxide reductase complex subunit 1* (*VKORC1*), which encodes a small transmembrane protein of the endoplasmic reticulum. *VKORC1* contains missense mutations in both human disorders and in a warfarin-resistant rat strain. Overexpression of wild-type *VKORC1*, but not *VKORC1* carrying the VKCFD2 mutation, leads to a marked increase in VKOR activity, which is sensitive to warfarin inhibition.

We previously mapped the locus for VKCFD2 to a 20-megabase (Mb) region on chromosome 16p12–q21 (ref. 9). We noted that a linkage group of genes on 16p11 is orthologous to genes around the warfarin resistance loci *Rw* in rats^{10,11} and *War* in mice¹². This homology led us to propose that VKCFD2 and warfarin resistance may be caused by allelic mutations in the same gene. If so, this would narrow down the critical interval in humans to a region of roughly 4.0 Mb on the short arm of chromosome 16. According to the human genome assembly (build 33, April 2003), this region contains 129 Ensembl genes with roughly 1,100 exons (<http://www.ensembl.org/>).

Using genomic DNA from two VKCFD2 and four WR probands, we carried out a systematic mutation screen by directly sequencing genes from this region and found missense mutations in a gene of

unknown function in all investigated VKCFD2 and WR probands (Fig. 1). This gene (IMAGE3455200) extends over a genomic region of 5,126 base pairs (bp) and comprises three exons encoding a protein of 163 amino acids with a calculated relative molecular mass of 18,000 (M_r 18K). Both non-related VKCFD2 index patients and their affected siblings carried the same homozygous point mutation in the third exon (292C $\rightarrow$ T), whereas the parents were found to be heterozygous (Fig. 1). The mutation caused the replacement of Arg 98 by tryptophan. The families were of Lebanese and German origin. Their haplotypes in the region of homozygosity encompassing the mutated gene were different. Furthermore, cytosine 292 is part of a CpG dinucleotide, which is a known mutation hotspot. Thus, the two mutations most probably arose independently.

Four different heterozygous mutations were detected in the individuals with WR, as expected for the autosomal dominant inheritance of this phenotype. Patient C showed a valine to leucine substitution (Val29Leu), patient D a valine to alanine substitution (V45A), patient E an arginine to glycine replacement (Arg58Gly), and patient F an exchange of leucine to arginine (Leu128Arg). The Arg58Gly mutation was shared by the two affected brothers of patient E. Sequencing the whole coding region in 384 control chromosomes detected no amino acid substitutions but two synonymous single nucleotide polymorphisms (129C $\rightarrow$ T, Cys43Cys; 358C $\rightarrow$ T, Leu120Leu).

Because we had thought that warfarin resistance in rats might be due to mutations in the orthologous gene, we also screened this gene in the rat strain that was used to map *Rw* to chromosome 1 (ref. 10). DNA samples of ten sensitive and ten resistant rats were sequenced. We detected a heterozygous missense mutation, Tyr139Cys (416A $\rightarrow$ G), in ten resistant rats and in one sensitive rat that was absent from the other nine sensitive animals. Warfarin resistance determination by the blood-clotting response test (see Methods) is known to have false-positive and false-negative error rates of at least 0.01 (ref. 13). Under these conditions, linkage between mutation and *Rw* phenotype was still significant with a log likelihood ratio (lod) score of 3.9 at a recombination rate of zero.

We also sequenced 16 resistant and 5 sensitive wild-caught rats¹⁴. The sensitive rats carried a homozygous tyrosine at position 139, whereas 12 and 4 resistant rats were homozygous and heterozygous, respectively, for the mutation Tyr139Cys. The identification of seven independent mutations with two distinct phenotypes in humans and rat, and the absence of mutations from 384 human control chromosomes indicated that mutations in this gene might have a causal role in the pathogenesis of VKCFD2 and warfarin resistance. On the basis of the mutations and the functional data reported below, we named the gene *vitamin K epoxide reductase complex subunit 1* (*VKORC1*).

An orthologue of the *VKORC1* gene was present in mice (NM_178600), and the orthologues in rat and in *Fugu rubripes* (Fig. 2) were established by homology searches and polymerase chain reaction with reverse transcription (RT-PCR). The corresponding proteins share 79–84% identity with the human protein. Database searches did not detect homology to any known genes or to characterized protein domains. Topology prediction programs anticipated three transmembrane domains. The PSORT II (<http://www.psort.org/>) program predicted a carboxy-terminal endoplasmic reticulum (ER) membrane retention signal (Lys-Lys-X-X or Lys-X-Lys-X-X) in all *VKORC1* proteins¹⁵, in accordance with location of the VKOR activity in the ER membrane fraction (ref. 5 and Fig. 3).

Tblastn searches with *VKORC1* detected paralogous human (BC027734) and mouse (AK009497) genes, each with 50% protein identity. We established the complete complementary DNA sequences in mouse, rat and *Fugu rubripes*. We named these genes *VKORC1-like 1* (*VKORC1L1*). Human, mouse and rat *VKORC1L1* proteins were highly conserved, sharing roughly 97% identity. We also detected a homologous protein in *Xenopus laevis* (AAH43742)

and, with weaker homology, in *Anopheles gambiae* (EAA06271) (Fig. 2). Pseudogenes of *VKORC1* and *VKORC1L* were present in the human, mouse and rat genomes.

VKORC1 seems to be conserved in vertebrates, as it is present in human, rodents and fish. Notably, a close homologue is present in *Anopheles gambiae* but not in the *Drosophila* genome, although a gene encoding γ -carboxylase has been described in the fruitfly¹⁶. γ -Carboxyglutamate residues have been also found in the peptide venom of the marine snail genus *Conus*¹⁷. Like in vertebrates, γ -carboxylation in the mollusc and *Drosophila* requires reduced vitamin K as a cofactor. Thus, protein modification by γ -carboxylation seems to antedate the evolutionary emergence of *VKORC1*.

The tissue distribution of more than 100 human expressed sequence tags available at UniGene suggested that *VKORC1* is widely expressed. On analysis of fetal and adult human tissues by northern blotting, we detected a single transcript of 1.0 kilobases, providing no evidence for alternative splicing. We found the highest expression in fetal and adult liver, followed by fetal heart, kidney and lung, adult heart and pancreas (data not shown). Thus, like the *GGCX* gene, *VKORC1* is highly expressed in the adult liver; however, its expression seems to be broader. For example, it is clearly transcribed in fetal liver and both fetal and adult heart, where only low expression of *GGCX* has been described¹⁸.

From biochemical fractionation experiments it is known that the VKOR activity purifies with the microsomal membrane fraction^{5,19}. Furthermore, γ -glutamyl carboxylase has been localized to the membranes of the ER by immunocytochemistry²⁰. To study the subcellular localization of human *VKORC1*, we generated constructs expressing green fluorescent protein (GFP)- and Myc-tagged *VKORC1* fusion proteins for transient transfection of COS-7 cells. Primary antibodies against the epitope tags and fluorochrome-labelled secondary antibodies were used to visualize the fusion proteins; an antibody against the ER-specific protein calnexin was

used as a control. The green immunofluorescence of the *VKORC1* fusion proteins decorated the mesh-like structures of the ER in the cytoplasm and perfectly colocalized with the red label of the ER marker calnexin (Fig. 3). Thus, *VKORC1* seems to be located in the ER.

To study the effect of *VKORC1* on the reduction of vitamin K 2,3-epoxide, we overexpressed wild-type and mutated *VKORC1* in HEK293 cells. VKOR activity was measured by the dithiothreitol (DTT)-dependent conversion of vitamin K 2,3-epoxide to vitamin K quinone. Dose-response curves to warfarin inhibition were measured at final concentrations of 5–80 μ M warfarin¹⁹. Untransfected and mock-transfected cells showed a low basal activity, which was sensitive to warfarin. Overexpression of wild-type *VKORC1* resulted in a marked stimulation of VKOR activity: production of vitamin K quinone was increased 14–21-fold as compared with untreated and mock-transfected cells. The activity was inhibited by warfarin in a dose-dependent manner (Table 1).

We also determined VKOR activity after transfection with constructs expressing mutated *VKORC1* (Table 1). Recombinant expression of the Arg98Trp mutation observed in the two VKCFD2 families only slightly increased VKOR activity in HEK293 cells. Spontaneous bleeding episodes and high concentrations of serum vitamin K epoxide in these patients suggest that the efficiency of vitamin K recycling is also markedly decreased *in vivo*²¹. The five WR variants showed a reduced VKOR activity ranging from 5% for the Leu128Arg mutation to 96% for the Val29Leu mutation. The Val45Ala, Arg58Gly and Tyr139Cys variants showed, respectively, about 23%, 21% and 48% activity (Table 1).

Reduced VKOR activity associated with higher vitamin K demand and death from spontaneous bleeding has been observed in heterozygous and homozygous *Rw* rats^{13,14,22}. Similarly in our expression system, which mimics homozygous conditions, WR

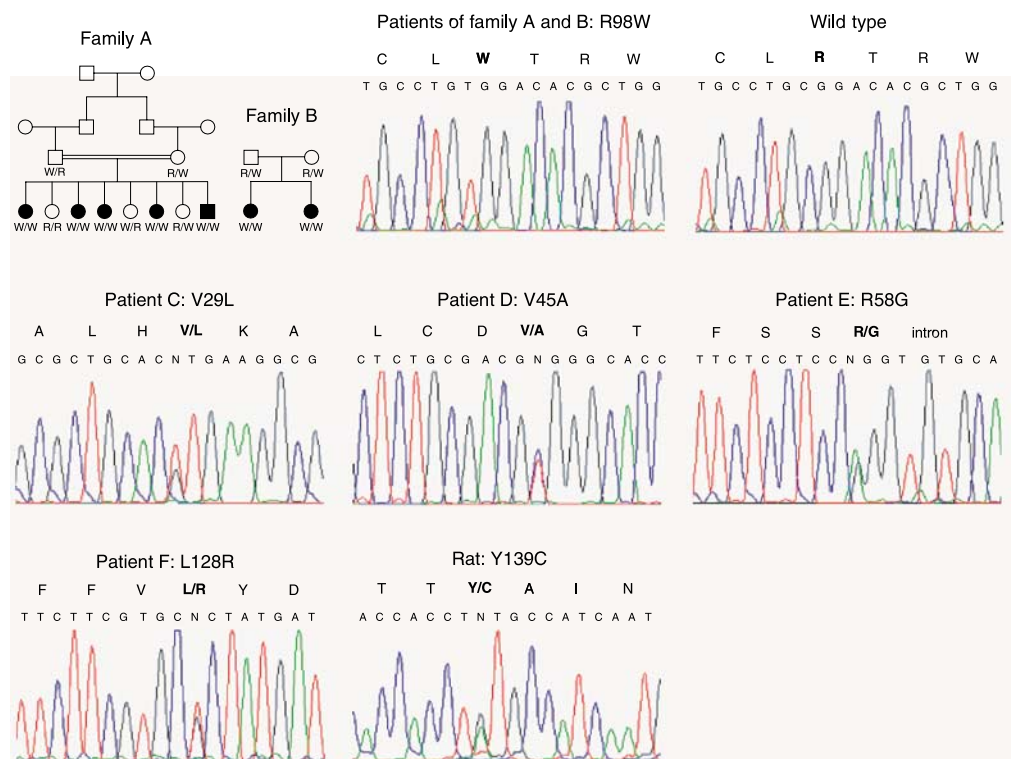


Figure 1 *VKORC1* mutations in individuals with VKCFD2 and WR. Top, segregation of the Arg98Trp mutation in two families with VKCFD2 (left), and the electropherograms for a patient with a homozygous mutation (centre) and a control subject (right). Middle and

bottom, the electropherograms for four WR patients with heterozygous mutations (85G $\rightarrow$ T, 134T $\rightarrow$ C, 172A $\rightarrow$ G, 383T $\rightarrow$ G) and a heterozygous *Rw* rat (416A $\rightarrow$ G).

mutations led to a lower functional efficiency of the VKOR complex. Whereas at the phenotypic level all WR variants showed at least partial resistance towards the anticoagulation effect of warfarin, both wild-type and mutant proteins were sensitive to warfarin *in vitro*. At concentrations above 20 μ M, the Val29Leu and

Tyr139Cys variants retained higher VKOR activities than did the wild type, whereas in the Val45Ala, Arg58Gly and Leu128Arg variants VKOR activity fell below the detection limit (Table 1). Inconsistent response to warfarin has been reported previously in warfarin-resistant rats^{14,22}. Further studies are required to resolve in

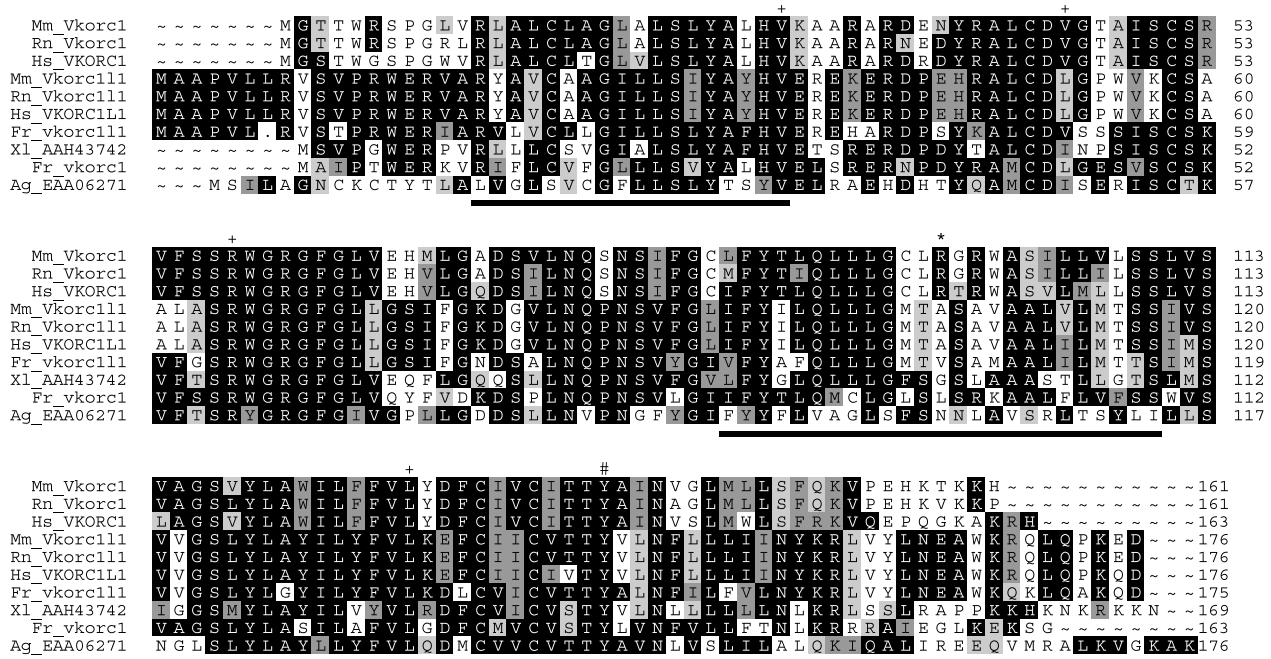


Figure 2 Amino acid sequence alignment of VKORC1 and VKORC1L1. The alignment was generated with ClustalW and Prettybox. Human, mouse and rat VKORC1 and VKORC1L1 proteins share roughly 84% and 97% identity, respectively, and there is about 50% identity between the two groups of proteins. Homologues were also detected in *Xenopus laevis* and *Anopheles gambiae*. The proteins are labelled by their gene symbols or accession codes and a prefix indicating the species (Hs, *H. sapiens*; Mm, *M. musculus*;

Rn, *R. norvegicus*; Fr, *Fugu rubripes*; Xl, *X. laevis*; Ag, *A. gambiae*). Predicted transmembrane domains are indicated by bold lines. Residues 29, 58 and 128 (+), which are mutated in individuals with WR, and residue 139 (#), which is mutated in *Rw* rats, are conserved in all species. WR mutation Val45Ala (+) and Arg 98 (*), which is mutated in individuals with VKCFD2, are conserved in human, rat and mouse VKORC1.

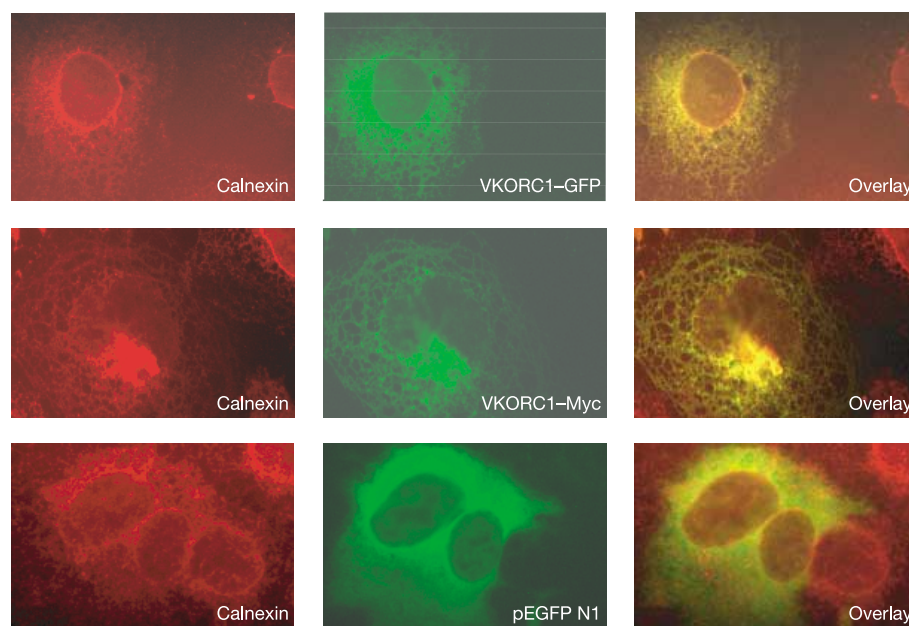


Figure 3 Subcellular location of VKORC1. COS-7 cells transiently transfected with constructs expressing VKORC1 were stained with anti-calnexin (red; left column) and anti-GFP or anti-Myc, respectively (green; middle column). Merged images of the double-

stained cells are shown on the right. Both VKORC1 constructs (tagged with GFP or Myc) colocalize with the ER-specific calnexin staining, whereas the control construct (pEGFP-N1) shows a diffuse pattern of staining throughout the cytoplasm.

Table 1 **VKOR activities* of HEK293 cells transfected with VKORC1 cDNA**

Mutation	Phenotype	Warfarin (μ M)					
		0	5	10	20	40	80
Wild type	Normal	20.29	8.66	6.39	4.33	2.71	0.88
Arg98Trp	VKCFD2	1.81	0.73	0.42	0.16	<0.10	<0.10
Val29Leu	WR	19.59	6.34	5.08	5.26	4.69	2.19
Val45Ala	WR	4.67	1.12	1.14	0.57	<0.10	<0.10
Arg58Gly	WR	4.20	1.02	1.03	0.72	0.40	<0.10
Leu128Arg	WR	1.05	0.30	0.31	0.31	<0.10	<0.10
Tyr139Cys	Rw	9.74	10.44	9.47	8.77	8.23	6.71

* Enzymatic activity was determined as described in Methods. Values are given as the percentage of vitamin K epoxide converted to vitamin K quinone (product/[substrate + product]). All tests were done in duplicate. Untransfected and mock-transfected HEK293 cells showed an activity of 1.49 and 0.96%, respectively, and were >90% inhibited by 10 μ M warfarin.

more detail the function of VKORC1 in warfarin resistance and VKCFD2.

In summary, by using a positional cloning approach that integrated mapping information from three species, we have identified a gene that is mutated in VKCFD2 and WR. From the segregation data, the absence of mutations in controls and the activity measurements of recombinant proteins, we conclude that mutations in this gene are causative for both VKCFD2 and WR. The molecular characterization of this first component of the VKOR complex will be instrumental in isolating the other subunits of this complex and in resolving the binding mechanisms of substrate and inhibitors. This should lead to a better understanding of the pharmacological action of coumarin drugs and may provide a basis for the rational design of anticoagulants targeting VKOR. □

Methods

Subjects

We carried out mutation screening of candidate genes on DNA samples from two VKCFD2 index patients and three unrelated WR patients. Clinical data of the VKCFD2 families have been described²¹. Patients C, D and F are sporadic cases. Patient E has two brothers suffering from warfarin resistance. Whereas the normal range of weekly warfarin doses is 10–60 mg, patient C required 100 mg, and patient E and his two brothers needed 220–250 mg of warfarin per week to reach therapeutic anticoagulation. Patients D and F did not respond to warfarin at all doses tested. All patients gave informed consent to the study.

Rw rats

An *Rw* rat colony was established as described¹⁰. A male brown Norway rat caught in the wild from the Münsterland area, Germany, that was homozygous resistant to warfarin and bromadiolone was crossed with a Wistar albino female. One F₁ heterozygous resistant male was used to establish a segregating backcross population. At the time of the study, the colony was abandoned. We analysed DNA samples from seven susceptible and eight resistant offspring of four matings between two sixth and seventh generation heterozygous resistant males and Wistar albino rats, and three susceptible and two resistant offspring of a backcross between a seventh generation heterozygous resistant male and his susceptible daughter from an earlier mating. Rats were tested for warfarin and bromadiolone resistance with the blood-clotting response method. Wild-caught rats were characterized for resistance to coumarin-derivatives and VKOR activity as described¹⁴.

Sequence analysis

Genome sequences and annotation were obtained from the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>), University of California Santa Cruz (UCSC) (<http://genome.ucsc.edu/>) and Ensembl (<http://www.ensembl.org/>). Primers for mutation screening were designed using Primer3 software integrated into a script to allow for automatic primer design (<http://ihg.gsfc.de/ihg/ExonPrimer.html>). We amplified exon sequences and exon–intron boundaries with intronic primers and sequenced them directly by the BigDye Terminator Cycle sequencing kit (ABI). Primer sequences are available from the authors on request. Topology predictions were done using PHDhtm (<http://cubic.bioc.columbia.edu/predictprotein/>).

Northern blots

Human multiple tissue northern blots, Fetal Blot 1 (Stratagene) and Human 8-Lane (BD Biosciences Clontech), were hybridized to radiolabelled full-length human VKORC1 cDNA in MiracleHyb high-performance hybridization solution (Stratagene). A β -actin probe supplied with the multiple tissue northern blot was used as a control.

Cloning and construction of expression vectors

The complete coding sequence of VKORC1 was amplified from human liver and kidney cDNA by using Marathon-Ready cDNA (BD Biosciences Clontech) and primers with added cleavage sites for HindIII and EcoRI (VKORC1-HindIII-F, 5'-ATTAAGCTTCACCA

TGGGCAGCACCTGGGGGAGCCCT-3'; VKORC1-EcoRI-R, 5'-ATTGAATTCGGTG CCTCTTAGCCTTGGCCCTG-3'), and cloned into pBluescript II (Stratagene). For immunocytochemistry, we re-cloned the insert into the mammalian expression vectors pEGFP-N1 (BD Biosciences Clontech) and pcDNA3.1/myc-His (Invitrogen).

For expression studies, the VKORC1 cDNA was cloned into pcDNA3 (Invitrogen) after amplification with the primers VKORC1-pcdna3-F (5'-GGGCGGAAGCTTGAGATAA TGGGCA-3') and VKORC1-pcdna3-R (5'-GCTTGAATTCAGGGCTCAGTGC-3'). Mutagenesis was done with the QuikChange mutagenesis kit (Stratagene). Wild-type and mutated cDNAs were re-cloned for expression in pCEP4 (Invitrogen) using the HindIII and XhoI sites. We verified all constructs by sequencing.

Cell culture, transient transfection and immunocytochemistry

COS-7 cells (DSMZ) were grown in DMEM medium with 10% fetal calf serum (FCS) on glass coverslips for 18–24 h and transfected with the expression vectors by using Effectene (Qiagen) in accordance with the manufacturer's specifications. After 48–60 h of further culturing, immunostaining was done as described²². Primary antibodies, Living Colors A.v. JL-8 (BD Biosciences Clontech), anti-Myc antibody (Invitrogen) and anti-calnexin (Sigma), were diluted 1:100 in the blocking solution. The same incubation and washing procedures were used for the secondary antibodies fluorescein isothiocyanate (FITC)-conjugated anti-mouse IgG (Sigma) and Cy3-conjugated anti-rabbit IgG F(ab')₂ fragment (Sigma). Coverslips were counterstained with 4',6-diamidino-2-phenylindole dihydrochloride (DAPI; 1:500), washed and visualized under a fluorescence microscope (Leica).

Expression studies and VKOR activity measurements

We grew HEK293-EBNA cells (Invitrogen) in MEM medium plus 10% FCS. For each experiment, 6 × 10⁵ cells were plated onto 94-mm Petri dishes. After 30 h at 37 °C and 5% CO₂, transfection (20 μ g of DNA construct per dish) was done by the calcium phosphate method. After 40 h at 35 °C and 3% CO₂, transfected cells were washed in PBS, collected and lysed in 0.25 mM imidazole (pH 7.6) plus 0.5% CHAPS. Transfection efficiency was checked by sequencing of RT-PCR products.

VKOR enzymatic activity was measured in whole-cell extracts as described¹⁹ with 5 μ M vitamin K 2,3-epoxide as a substrate in the presence of 5 mM DTT. After extraction by isopropanol:hexane (3:2), vitamin K quinone was separated from the epoxide by HPLC on a reversed-phase C₁₈ column. During the extraction procedure, vitamin K hydroquinone was quantitatively oxidized to the quinone form. Measurements were run in duplicate and the activity is given as the percentage of substrate converted into quinone. Vitamin K 2,3-epoxide was prepared by the oxidation of vitamin K quinone (Sigma-Aldrich) with H₂O₂. We added warfarin (Sigma-Aldrich) in dimethyl sulphoxide (<1 vol%).

Received 5 September; accepted 31 October 2003; doi:10.1038/nature02214.

- Suttie, J. W. The biochemical basis of warfarin therapy. *Adv. Exp. Med. Biol.* **214**, 3–16 (1987).
- Nelsetuen, G. L., Zytovicz, T. H. & Howard, J. B. The mode of action of vitamin K. Identification of γ -carboxylglutamic acid as a component of prothrombin. *J. Biol. Chem.* **249**, 6347–6350 (1974).
- Stenflo, J., Fernlund, P., Egan, W. & Roepstorff, P. Vitamin K dependent modifications of glutamic acid residues in prothrombin. *Proc. Natl Acad. Sci. USA* **71**, 2730–2733 (1974).
- Fasco, M. J., Principe, L. M., Walsh, W. A. & Friedman, P. A. Warfarin inhibition of vitamin K 2,3-epoxide reductase in rat liver microsomes. *Biochemistry* **22**, 5655–5660 (1983).
- Cain, D., Hutson, S. M. & Wallin, R. Assembly of the warfarin-sensitive vitamin K 2,3-epoxide reductase enzyme complex in the endoplasmic reticulum membrane. *J. Biol. Chem.* **272**, 29068–29075 (1997).
- Begent, L. A. *et al.* Characterization and purification of the vitamin K₁ 2,3-epoxide reductases system from rat liver. *J. Pharm. Pharmacol.* **53**, 481–486 (2001).
- Lee, J. J. & Fasco, M. J. Metabolism of vitamin K and vitamin K 2,3-epoxide via interaction with a common disulfide. *Biochemistry* **23**, 2246–2252 (1984).
- Wallin, R., Hutson, S. M., Cain, D., Sweatt, A. & Sane, D. C. A molecular mechanism for genetic warfarin resistance in the rat. *FASEB J.* **15**, 2542–2544 (2001).
- Fregin, A. *et al.* Homozygosity mapping of a second gene locus for hereditary combined deficiency of vitamin K-dependent clotting factors to the centromeric region of chromosome 16. *Blood* **100**, 3229–3232 (2002).
- Kohn, M. H. & Pelz, H. J. Genomic assignment of the warfarin resistance locus, *Rw*, in the rat. *Mamm. Genome* **10**, 696–698 (1999).
- Greaves, J. H. & Ayres, P. Heritable resistance to warfarin in rats. *Nature* **215**, 877–878 (1967).
- Wallace, M. E. & MacSwiney, F. J. A major gene controlling warfarin-resistance in the house mouse. *J. Hyg. (Lond.)* **76**, 173–181 (1976).
- Martin, A. D., Steed, L. C., Redfern, R., Gill, J. E. & Huson, L. W. Warfarin-resistance genotype determination in the Norway rat. *Rattus norvegicus. Lab. Anim.* **13**, 209–214 (1979).
- Thijssen, H. H. & Pelz, H. J. in *Advances in Vertebrate Pest Management* (eds Pelz, H. J., Cowan, D. P. & Feare, C. J.) 181–192 (Filander, Fürth, 2001).
- Jackson, M. R., Nilsson, T. & Peterson, P. A. Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. *EMBO J.* **9**, 3153–3162 (1990).
- Li, T., Yang, C. T., Jin, D. & Stafford, D. W. Identification of a *Drosophila* vitamin K-dependent γ -glutamyl carboxylase. *J. Biol. Chem.* **275**, 18291–18296 (2000).
- Bandyopadhyay, P. K. *et al.* Gamma-glutamyl carboxylation: an extracellular posttranslational modification that antedates the divergence of molluscs, arthropods, and chordates. *Proc. Natl Acad. Sci. USA* **99**, 1264–1269 (2002).
- Romero, E. E., Velazquez-Estades, L. J., Deo, R., Schapiro, B. & Roth, D. A. Cloning of rat vitamin K-dependent γ -glutamyl carboxylase and developmentally regulated gene expression in postimplantation embryos. *Exp. Cell Res.* **243**, 334–346 (1998).
- Wallin, R. & Martin, L. F. Vitamin K-dependent carboxylation and vitamin K metabolism in liver. Effects of warfarin. *J. Clin. Invest.* **76**, 1879–1884 (1985).
- Presnell, S. R. & Stafford, D. W. The vitamin K-dependent carboxylase. *Thromb. Haemost.* **87**, 937–946 (2002).
- Oldenburg, J. *et al.* Congenital deficiency of vitamin K dependent coagulation factors in two families

presents as a genetic defect of the vitamin K-epoxide- reductase complex. *Thromb. Haemost.* **84**, 937–941 (2000).

22. Fasco, M. J., Preusch, P. C., Hildebrandt, E. & Suttie, J. W. Formation of hydroxyvitamin K by vitamin K epoxide reductase of warfarin-resistant rats. *J. Biol. Chem.* **258**, 4372–4380 (1983).
23. Hörtnagl, K., Prokisch, H. & Meitinger, T. An isoform of hPANK2, deficient in pantothenate kinase-associated neurodegeneration, localizes to mitochondria. *Hum. Mol. Genet.* **12**, 321–327 (2003).

Acknowledgements We thank H. Hermann-Brackmann, W. Eberl, J. Pattinson, A.-N. Parkes and R. Jurgutis for the donation and clinical characterization of patient samples; V. Milenkovic for technical assistance and H. Höhn, T. Meitinger and T. Wienker for discussions and critically reading the manuscript. This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG), the Bundesministerium für Bildung und Forschung Deutsches Zentrum für Luft- und Raumfahrt (BMBF/DLR), Baxter Germany, the Stiftung Hämostherapie-Forschung, the Gesellschaft für Thrombose- und Hämostaseforschung (GTH) and the BMBF projects German National Genome Research Network (NGFN) and Bioinformatics for the Functional Analysis of Mammalian Genomes (BFAM).

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for material should be addressed to J.O. (joldenburg@bsd.hessen.de). The sequences are deposited in GenBank under accession codes AY423042, AY423043, AY423044, AY423045, AY423046, AY423047 and AY423048

Identification of the gene for vitamin K epoxide reductase

Tao Li¹, Chun-Yun Chang¹, Da-Yun Jin¹, Pen-Jen Lin¹, Anastasia Khvorova² & Darrel W. Stafford¹

¹Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA

²Dharmacon, Inc., 1376 Miners Drive 101, Lafayette, Colorado 80026, USA

Vitamin K epoxide reductase (VKOR) is the target of warfarin, the most widely prescribed anticoagulant for thromboembolic disorders. Although estimated to prevent twenty strokes per induced bleeding episode¹, warfarin is under-used because of the difficulty of controlling dosage and the fear of inducing bleeding. Although identified in 1974 (ref. 2), the enzyme has yet to be purified or its gene identified. A positional cloning approach has become possible after the mapping of warfarin resistance to rat chromosome 1 (ref. 3) and of vitamin K-dependent protein deficiencies to the syntenic region of human chromosome 16 (ref. 4). Localization of VKOR to 190 genes within human chromosome 16p12-q21 narrowed the search to 13 genes encoding candidate transmembrane proteins, and we used short interfering RNA (siRNA) pools against individual genes to test their ability to inhibit VKOR activity in human cells. Here, we report the identification of the gene for VKOR based on specific inhibition of VKOR activity by a single siRNA pool. We confirmed that MGC11276 messenger RNA encodes VKOR through its expression in insect cells and sensitivity to warfarin. The expressed enzyme is 163 amino acids long, with at least one transmembrane domain. Identification of the VKOR gene extends our understanding of blood clotting, and should facilitate development of new anticoagulant drugs.

Numerous proteins require the modification of multiple glutamic acid residues to γ -carboxyglutamate in order to function. Among these vitamin-K-dependent (VKD) proteins, coagulation factors IX (Christmas factor), VII and prothrombin are the best known. The observation that a knockout of the gene encoding matrix Gla protein results in calcification of arteries in the mouse³ emphasizes the importance of the vitamin K cycle for proteins with functions other than coagulation. Moreover, Gas6 (ref. 6) and other Gla proteins of unknown function^{7,8} are expressed in neural tissue, and warfarin exposure *in utero* results in mental retardation and

facial abnormalities^{9–11}. This is consistent with the observation that the expression of VKD γ -glutamyl carboxylase, the enzyme that accomplishes the Gla modification^{12,13}, is temporally regulated in a tissue-specific manner with high expression in the nervous system during early embryonic stages¹⁴. Concomitant with carboxylation, reduced vitamin K, a co-substrate of the reaction, is converted to vitamin K epoxide. Because the amount of vitamin K in the human diet is limited, vitamin K epoxide must be converted back to vitamin K by VKOR to prevent its depletion. Warfarin, the most widely used anticoagulation drug, targets VKOR and prevents the regeneration of vitamin K. The consequence is a decrease in the concentration of reduced vitamin K, which results in a decreased rate of carboxylation by the γ -glutamyl carboxylase and in the production of under-carboxylated VKD proteins.

VKOR activity was first reported in 1974 (ref. 2). Since this first report, several groups, including ours, have attempted to purify VKOR with little success. Two recent reports suggesting the chromosomal location of the gene for VKOR stimulated us to attempt a different approach to identify the gene for this important enzyme. First, Kohn and Pelz³ mapped warfarin resistance (*R_w*) to rat chromosome 1 between markers *D1Rat193* and *D1Rat130*. In this region, four marker genes (*Il4r*, *Myl2*, *Itgam* and *Fgfr2*) allowed Kohn and Pelz to locate the syntenic regions of mouse and human chromosomes³. They reported that these four markers are found on mouse chromosome 7 at 60–63 centimorgans (cM) and on human chromosomes 10, 12 and 16. A second study⁴ mapped a gene for combined deficiencies of VKD proteins to human chromosome 16p12-q21, between markers *D16S3131* and *D16S419*. We found human genes corresponding to three of the four rat marker genes (*Il4r*, *Myl2* (rat *Myl2* is 97% identical to human *HUMMLC2B*) and *Itgam*) on human chromosome 16p12-q21. Because warfarin resistance and combined deficiency of VKD proteins map to the same region, we focused our search on human chromosome 16. This region between markers *D16S3131* and *D16S419* corresponds to chromosome 16 at 50–65 cM on the genetic map and to 26–46.3 megabases (Mb) on the physical map (<http://www.ncbi.nih.gov/Genomes/>). We analysed the 190 predicted coding sequences in this region by a BLASTX search of the NCBI non-redundant protein database. Those human genes and orthologues from related species with known function were eliminated. Because VKOR seems to be a transmembrane protein¹⁵, the remaining 52 genes were translated

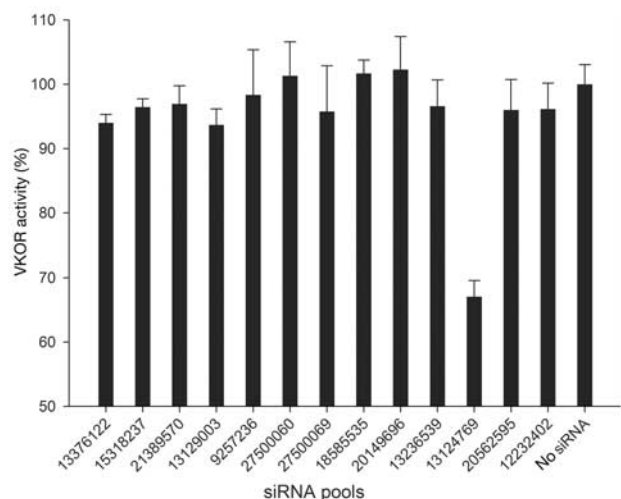


Figure 1 siRNA inhibition of VKOR activity in A549 cells. For each of the 13 siRNA pools, three T75 flasks containing A549 cells were transfected and VKOR activity was determined after 72 h. The VKOR assay used 25 μ M vitamin K epoxide. One siRNA pool specific for gene gi:13124769 reduced VKOR activity by 30–36% in eight repetitions.

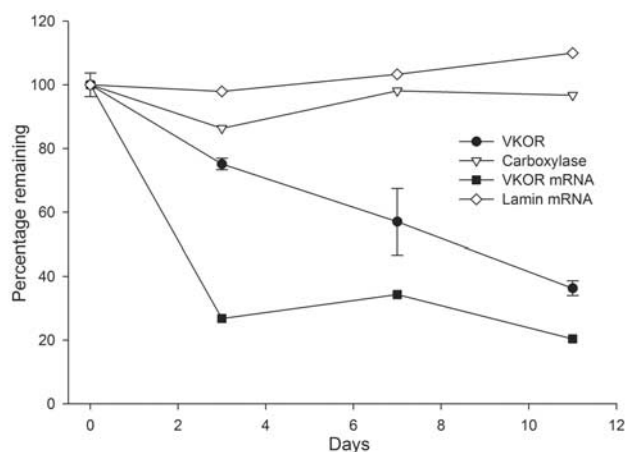


Figure 2 Time course of the inhibition of VKOR activity by the siRNA pool specific for gi:13124769 in A549 cells. VKOR activity decreased continuously during this time period while its mRNA levels decreased rapidly to about 20% of normal. A 25 μ M concentration of vitamin K epoxide was used for this assay. The siRNA does not affect the activity of VKD carboxylase or the level of lamin A/C mRNA.

according to the complementary DNA sequences in the NCBI database and analysed with the programs TMHMM and TMAP (<http://workbench.sdsc.edu/>) to predict those with transmembrane domains. Thirteen genes predicted to code for integral membrane proteins were chosen for further analysis.

Our strategy was to identify a cell line expressing relatively high amounts of VKOR activity and to use siRNA (double-stranded RNA of 21–23 nucleotides) to systematically knockdown all 13 candidate genes. siRNA causes specific RNA degradation in cell culture^{16–18}; however, application of siRNA for large-scale screening in mammalian cells has not been reported previously because of the difficulty in identifying a functional target for a specific mammalian cell mRNA¹⁹. The recent development of a rational selection algorithm²⁰ for siRNA design increases the probability that a specific siRNA can be developed; furthermore, we increased the probability of success by pooling four rationally selected siRNAs. Using siRNA to search for previously unidentified genes has the advantage that, even if VKOR activity requires the product of more than one gene for activity, the screen should still be effective because the assay determines the loss of enzymatic activity.

We screened 15 cell lines and found a human lung carcinoma cell line, A549, which exhibited sufficient warfarin-sensitive VKOR activity for easy measurement. A second human colorectal adenocarcinoma cell line (HT-29) that expressed very little VKOR activity was used as a reference.

We transfected each of the 13 pools of siRNA in triplicate into A549 cells and after 72 h assayed for VKOR activity. One siRNA pool specific for a particular gene (referred to hereafter by its GenBank identification number gi:13124769) reduced VKOR activity by 30–36% in eight separate assays (Fig. 1). Inhibition with individual siRNAs from the gi:13124769 siRNA pool was also done. siRNAs 1 and 2 (see Methods for sequences) decreased VKOR activity by about 10% and 25%, respectively, and mRNA levels by 58% and 72%. siRNAs 3 and 4 had no apparent effect on VKOR activity, but siRNA 4 caused a 36% reduction in mRNA levels (data not shown).

One possible reason that VKOR activity was only inhibited by about 35% of its initial activity after 72 h in the presence of the pooled siRNA is that the half-life of mammalian proteins varies greatly (from minutes to days)^{21–23}, and also because we are inhibiting mRNA and not enzyme activity. Therefore, we carried the cells through 11 days and monitored their VKOR activity. Figure 2 shows that gi:13124769 mRNA levels decreased rapidly

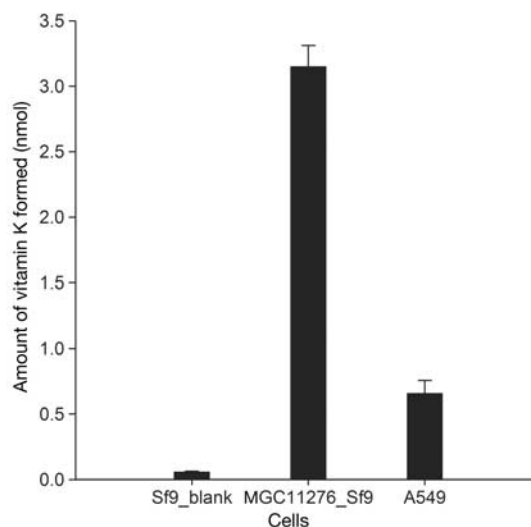


Figure 3 Expression of VKOR in Sf9 cells. VKOR activity was detected when MGC11276 was expressed in Sf9 insect cells. Approximately 1×10^6 cells were used in this assay. Reactions were performed using 32 μ M KO at 30 °C for 30 min in buffer D. Blank Sf9 cells served as a negative control and A549 cells served as a reference.

to about 20% of normal levels while VKOR activity decreased continuously during this time period. This reduction in activity is neither a nonspecific effect of the siRNA nor a result of cell death, because the levels of VKD carboxylase activity and of lamin A/C mRNA remained constant. Furthermore, the level of gi:13124769 mRNA is fourfold lower in HT-29 cells, which have low VKOR activity, than in A549 cells, which exhibit high VKOR activity (data not shown). These data suggest that gi:13124769 is the VKOR gene.

There are 338 cDNA clones in the NCBI database supporting the gene gi:13124769 and representing seven different splicing patterns (NCBI AceView program). These are composed of all or part of two to four exons. They have no other obvious features that would suggest function except that the isoform with four exons is predicted to localize to the mitochondrion. The most prevalent isoform, MGC11276, has three exons, and its sequence is supported by 158 cDNA clones in the database. Multiple clones representing the putative VKOR cDNA have been reported from almost every tissue, with the greatest number from lung, pancreas, nerve and liver. This three-exon transcript encodes a predicted protein of 163 amino acids with a mass of 18.2 kDa. It is a putative endoplasmic reticulum protein with one to three transmembrane domains, depending on the program used for prediction.

To confirm that we had identified the VKOR gene, we expressed the most prevalent form of the enzyme (three exons) in *Spodoptera frugiperda* (Sf9) cells. Sf9 cells have no measurable VKOR activity but exhibit VKOR activity when transfected with MGC11276 cDNA (Fig. 3). VKOR activity is observed in constructs having an epitope tag at either the amino or carboxy terminus. This tag should assist in the purification of VKOR.

Consistent with previous data³, the VKOR gene, IMAGE 3455200 (gi:13124769), maps to human chromosome 16p11.2, mouse chromosome 7F3 and rat chromosome 1 between *Myl2* and *Itgam*. The protein encoded by the VKOR gene has no known counterpart in the database to provide an indication of its function or structure. There are two DNA segments on the X chromosome and chromosome 1 that are 93% and 91% identical to the DNA coding for VKOR. These appear to be pseudogenes and neither DNA sequence is expected to code for a protein similar to VKOR.

VKOR has seven cysteine residues, consistent with observations that the enzymatic activity is dependent on thiol reagents²⁴. Five of the seven cysteines are conserved among human, mouse, rat,

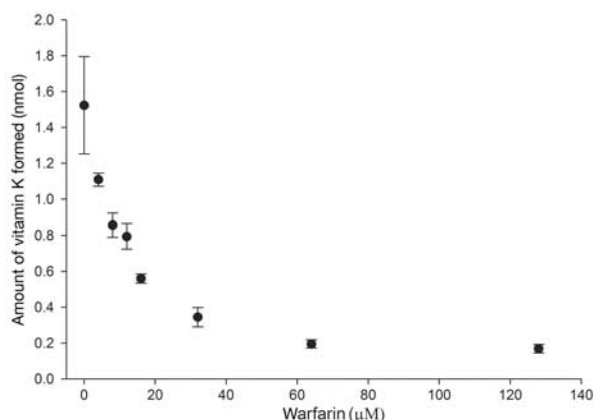


Figure 4 Inhibition of VKOR by warfarin. Reactions were performed using 1.6 mg microsomal proteins made from VKOR_Sf9 cells, 60 μM KO and various concentrations of warfarin at 30 °C for 15 min in buffer D.

zebrafish, *Xenopus* and *Anopheles*.

VKOR should exhibit warfarin sensitivity; therefore, we made microsomes from Sf9 cells expressing VKOR and tested for warfarin sensitivity. These tests revealed that VKOR activity is warfarin-sensitive (Fig. 4). Our observation that substantial activity results from the expression of VKOR in Sf9 insect cells suggests that the activity is the product of a single gene. However, we cannot yet rule out the possibility that other subunits or required cofactors are present in insect cells and act in conjunction with the expressed gene.

We report here the first example, to our knowledge, of using siRNA in mammalian cells to identify an unknown gene. The identity of the VKOR gene was confirmed by its expression in insect cells. The VKOR gene encodes several isoforms, and it will be important to characterize the activity and expression pattern of each isoform. Millions of people worldwide use warfarin to inhibit coagulation; therefore, it is important to characterize further VKOR as it may lead to more accurate dosing or to the design of safer, more effective anticoagulants than are currently available. □

Methods

siRNA design and synthesis

siRNAs were selected using an advanced version of a rational design algorithm²⁰. For each of the 13 genes, four siRNA duplexes predicted to have the greatest probability of success were selected and a BLAST search was conducted using the Human EST database. To minimize the potential for off-target silencing effects, only those sequence targets with more than three mismatches against unrelated sequences were selected²⁵. All duplexes were synthesized in Dharmacon as 21-nucleotides with UU overhangs using a modified method of 2'-ACE chemistry²⁶, and the anti-sense strand was chemically phosphorylated to ensure maximum activity²⁷. The siRNAs in the pool were designed to target the following nucleotides within gene gi:13124760: (1) CATATTCGGTTGCATCTTC; (2) ACACACTACAGCTATTGTT; (3) TCTACGCGCTGCACGTGAA; (4) TGTATCACCACCTATGCTA.

siRNA transfection

Transfection was essentially as described previously²⁸ with minor modifications. A final concentration of 100 nM for the siRNAs was used in the siRNA inhibition experiments in A549 cells. Transfection was carried out using TransIT-TKO according to the manufacturer's protocol (Mirus). Cells were usually assayed 72 h after transfection.

VKOR activity assay

SiRNA-transfected A549 cells were trypsinized and washed twice with cold PBS. 1.5×10^7 cells were taken for each VKOR assay. A total of 200 μl buffer D (250 mM Na₂HPO₄·NaH₂PO₄, 500 mM KCl, 20% glycerol and 0.75% CHAPS, pH 7.4) was added to the cell pellet, followed by sonication of the cell lysate. For assays of solubilized microsomes, microsomes were prepared from 2×10^9 transfected Sf9 cells as described²⁹, and 10–50 μl of solubilized microsomes were used for each assay. Vitamin K epoxide was added at the concentrations indicated in the figure legends, and dithiothreitol was added at 4 mM to initiate the reaction. The reaction mixture was incubated in yellow (or no) light at 30 °C for 30 min and was stopped by adding 500 μl of 0.05 M AgNO₃:isopropanol (5:9). A total of 500 μl hexane was added and the mixture was vortexed vigorously for 1 min to extract

the vitamin K and KO. Vitamin E was used as a reference standard. After 5 min centrifugation, the upper organic layer was transferred to a 5-ml brown vial and dried with N₂. A total of 150 μl high-performance liquid chromatography (HPLC) buffer, acetonitrile:isopropanol:water (100:7:2), was added to dissolve the vitamin K and KO, and the sample was analysed by HPLC on a C-18 column (Vydac; catalogue number 218TP54).

Quantitative PCR with reverse transcription

1×10^6 cells were washed with PBS twice, and total RNA was isolated with Trizol reagent according to the manufacturer's protocol (Invitrogen). A total of 1 μg RNA was digested by RQ1 DNaseI (Promega) and heat-inactivated. First strand cDNA was made with M-MLV reverse transcriptase (Invitrogen). cDNAs were mixed with DyNAmo SYBR Green qPCR pre-mix (Finnzymes), and real-time PCR was performed with an Opticon II PCR thermal cycler (MJ Research). We used the following primers: 13124769-5' (F), (TCCAACAGCATATTCGGTTGC); 13124769-3' (R), (TTCTTGACCTTCGGAAACT); GAPDH (F), (GAAGGTGAAGGTGCGAGTC); GAPDH (R), (GAAGATGGTGATGGGAT TTC); Lamin-RT (F), (CTAGGTGAGGCCAAGAAGCAA); Lamin-RT (R), (CTGTTCCTCTCAGCAGACTGC).

Expression of VKOR in Sf9 cell culture

The cDNA for the MGC11276 coding region was cloned into pVL1392 (Pharmingen) with the HPC4 tag (EDQVDPRLIDGK) at its N terminus, and was expressed in Sf9 cells as described³⁰.

Received 19 September; accepted 27 November 2003; doi:10.1038/nature02254.

- Horton, J. & Bushwick, B. Warfarin therapy: Evolving strategies in anticoagulation. *Am. Family Physician* **59**, 635–647 (1999).
- Zimmerman, A. & Matschiner, J. T. Biochemical basis of hereditary resistance to warfarin in the rat. *Biochem. Pharmacol.* **23**, 1033–1040 (1974).
- Kohn, M. H. & Pelz, H. J. A gene-anchored map position of the rat warfarin-resistance locus, *Rw*, and its orthologs in mice and humans. *Blood* **96**, 1996–1998 (2000).
- Fregin, A. *et al.* Homozygosity mapping of a second gene locus for hereditary combined deficiency of vitamin K-dependent clotting factors to the centromeric region of chromosome 16. *Blood* **100**, 3229–3232 (2002).
- Luo, G. *et al.* Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein. *Nature* **386**, 78–81 (1997).
- Manfioletti, G., Brancolini, C., Avanzi, G. & Schneider, C. The protein encoded by a growth arrest-specific gene (*gas6*) is a new member of the vitamin K-dependent proteins related to protein S, a negative coregulator in the blood coagulation cascade. *Mol. Cell. Biol.* **13**, 4976–4985 (1993).
- Kulman, J. D., Harris, J. E., Haldeman, B. A. & Davie, E. W. Primary structure and tissue distribution of two novel proline-rich γ-carboxyglutamic acid proteins. *Proc. Natl Acad. Sci. USA* **94**, 9058–9062 (1997).
- Kulman, J. D., Harris, J. E., Xie, L. & Davie, E. W. Identification of two novel transmembrane γ-carboxyglutamic acid proteins expressed broadly in fetal and adult tissues. *Proc. Natl Acad. Sci. USA* **98**, 1370–1375 (2001).
- Stevenson, R. E., Burton, O. M., Ferlauto, G. J. & Taylor, H. A. Hazards of oral anticoagulants during pregnancy. *J. Am. Med. Assoc.* **243**, 1549–1551 (1980).
- Barr, M. Jr & Burdi, A. R. Warfarin-associated embryopathy in a 17-week-old abortion. *Teratology* **14**, 129–134 (1976).
- Howe, A. M. & Webster, W. S. The warfarin embryopathy: a rat model showing maxillofacial hypoplasia and other skeletal disturbances. *Teratology* **46**, 379–390 (1992).
- Wu, S. M., Cheung, W. F., Frazier, D. & Stafford, D. W. Cloning and expression of the cDNA for human γ-glutamyl carboxylase. *Science* **254**, 1634–1636 (1991).
- Presnell, S. R. & Stafford, D. W. The vitamin K-dependent carboxylase. *Thromb. Haemost.* **87**, 937–946 (2002).
- Tsaioun, K. I. Vitamin K-dependent proteins in the developing and aging nervous system. *Nutr. Rev.* **57**, 231–240 (1999).
- Carlisle, T. L. & Suttie, J. W. Vitamin K dependent carboxylase: subcellular location of the carboxylase and enzymes involved in vitamin K metabolism in rat liver. *Biochemistry* **19**, 1161–1167 (1980).
- Hara, K. *et al.* Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action. *Cell* **110**, 177–189 (2002).
- Krichevsky, A. M. & Kosik, K. S. RNAi functions in cultured mammalian neurons. *Proc. Natl Acad. Sci. USA* **99**, 11926–11929 (2002).
- Burns, T. F., Fei, P., Scata, K. A., Dicker, D. T. & El-Deiry, W. S. Silencing of the novel p53 target gene *Snk/Plk2* leads to mitotic catastrophe in paclitaxel (Taxol)-exposed cells. *Mol. Cell. Biol.* **23**, 5556–5571 (2003).
- Holen, T., Amarzougui, M., Babaie, E. & Prydz, H. Similar behaviour of single-strand and double-strand siRNAs suggests they act through a common RNAi pathway. *Nucleic Acids Res.* **31**, 2401–2407 (2003).
- Khorova, A., Reynolds, A. & Jayasena, S. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 1–20 (2003).
- Zhang, W., Lane, R. D. & Mellgren, R. L. The major calpain isoforms are long-lived proteins. Design of an antisense strategy for calpain depletion in cultured cells. *J. Biol. Chem.* **271**, 18825–18830 (1996).
- Bohley, P. Surface hydrophobicity and intracellular degradation of proteins. *Biol. Chem.* **377**, 425–435 (1996).
- Dice, J. F. & Goldberg, A. L. Relationship between *in vivo* degradative rates and isoelectric points of proteins. *Proc. Natl Acad. Sci. USA* **72**, 3893–3897 (1975).
- Thijssen, H. H., Janssen, Y. P. & Vervoort, L. T. Microsomal lipamide reductase provides vitamin K epoxide reductase with reducing equivalents. *Biochem. J.* **297**, 277–280 (1994).
- Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21**, 635–637 (2003).
- Scaringe, S. A. Advanced 5'-silyl-2'-orthoester approach to RNA oligonucleotide synthesis. *Methods Enzymol.* **317**, 3–18 (2000).

27. Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
28. Harborth, J., Elbashir, S. M., Bechert, K., Tuschl, T. & Weber, K. Identification of essential genes in cultured mammalian cells using small interfering RNAs. *J. Cell Sci.* **114**, 4557–4565 (2001).
29. Lin, P. J. *et al.* The putative vitamin K-dependent γ -glutamyl carboxylase internal propeptide appears to be the propeptide binding site. *J. Biol. Chem.* **277**, 28584–28591 (2002).
30. Li, T., Yang, C. T., Jin, D. & Stafford, D. W. Identification of a *Drosophila* vitamin K-dependent γ -glutamyl carboxylase. *J. Biol. Chem.* **275**, 18291–18296 (2000).

Acknowledgements We are grateful to T. Vision for help with bioinformatics, R. Wallin for technical assistance with the VKOR assay, and S. Makarov for assistance with quantitative PCR. For editorial assistance we thank L. Gabiger. This work was supported by grants from the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.W.S. (dws@email.unc.edu).

Multiple transport modes of the cardiac $\text{Na}^+/\text{Ca}^{2+}$ exchanger

Tong Mook Kang¹ & Donald W. Hilgemann²

¹Department of Physiology, Center for Molecular Medicine, Samsung Biomedical Research Institute, Sungkyunkwan University School of Medicine, Suwon 440-746, Korea

²Department of Physiology, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9040, USA

The cardiac $\text{Na}^+/\text{Ca}^{2+}$ exchanger¹ (NCX1; ref. 2) is a bi-directional Ca^{2+} transporter that contributes to the electrical activity of the heart^{3,4}. When, and if, Ca^{2+} is exported or imported depends on the $\text{Na}^+/\text{Ca}^{2+}$ exchange ratio⁵. Whereas a ratio of 3:1 ($\text{Na}^+:\text{Ca}^{2+}$) has been indicated by Ca^{2+} flux equilibrium studies⁶, a ratio closer to 4:1 has been indicated by exchange current reversal potentials^{7,8}. Here we show, using an ion-selective electrode technique⁹ to quantify ion fluxes in giant patches¹⁰, that ion flux ratios are approximately 3.2 for maximal transport in either direction. With Na^+ and Ca^{2+} on both sides of the membrane, net current and Ca^{2+} flux can reverse at different membrane potentials, and inward current can be generated in the absence of cytoplasmic Ca^{2+} , but not Na^+ . We propose that NCX1 can transport not only 1 Ca^{2+} or 3 Na^+ ions, but also 1 Ca^{2+} with 1 Na^+ ion at a low rate. Therefore, in addition to the major 3:1 transport mode, import of 1 Na^+ with 1 Ca^{2+} defines a Na^+ -conducting mode that exports 1 Ca^{2+} , and an electroneutral Ca^{2+} influx mode that exports 3 Na^+ . The two minor transport modes can potentially determine resting free Ca^{2+} and background inward current in heart.

NCX1 transport ratios greater than 3:1 ($\text{Na}^+:\text{Ca}^{2+}$)^{7,8} might reflect Na^+ movements by means of channel-like activity, as in neurotransmitter transporters¹¹, or through uncoupled transport reactions, as in $\text{Na}^+/\text{glucose}$ carriers¹². Therefore, we first verified in giant patches from guinea-pig myocytes and NCX1-expressing BHK cells¹³ that NCX1 mediates no detectable conductance when Na^+ is present without Ca^{2+} , and vice versa^{14,15}, nor when Na^+ and Ca^{2+} are present on one side but not on the other side.

Our method to quantify ion flux⁹ is based on Fick's law: ion flux generates an ion gradient that is proportional to the flux. Forward NCX1 transport moves Ca^{2+} from the cytoplasmic side (referred to as Ca_i) of the inside-out patch into the pipette (referred to as Ca_o) with opposite movement of Na^+ generating an inward exchange current. In Fig. 1a we describe the Ca_o gradient occurring within the patch pipette when an inward exchange current is activated in a

BHK patch. To do so, the pipette contains 100 mM Na_o with free Ca_o adjusted carefully to 8 μM by comparison to buffered standards. When 0.1 mM free Ca_i is applied, the current (–12 pA) activates immediately. A Ca^{2+} ion-selective electrode (ISE) with its tip (2 μm) positioned 6 μm from the pipette orifice shows a 17 mV response over 1 min that corresponds to a 32 μM accumulation of Ca_o . The Ca_o gradient is analysed by moving the ISE with a micromanipulator between positions that are 6, 25, 40, 55 and 200 μm from the pipette orifice and have inner diameters (d_{1-5}) of 10.8, 22, 32, 38

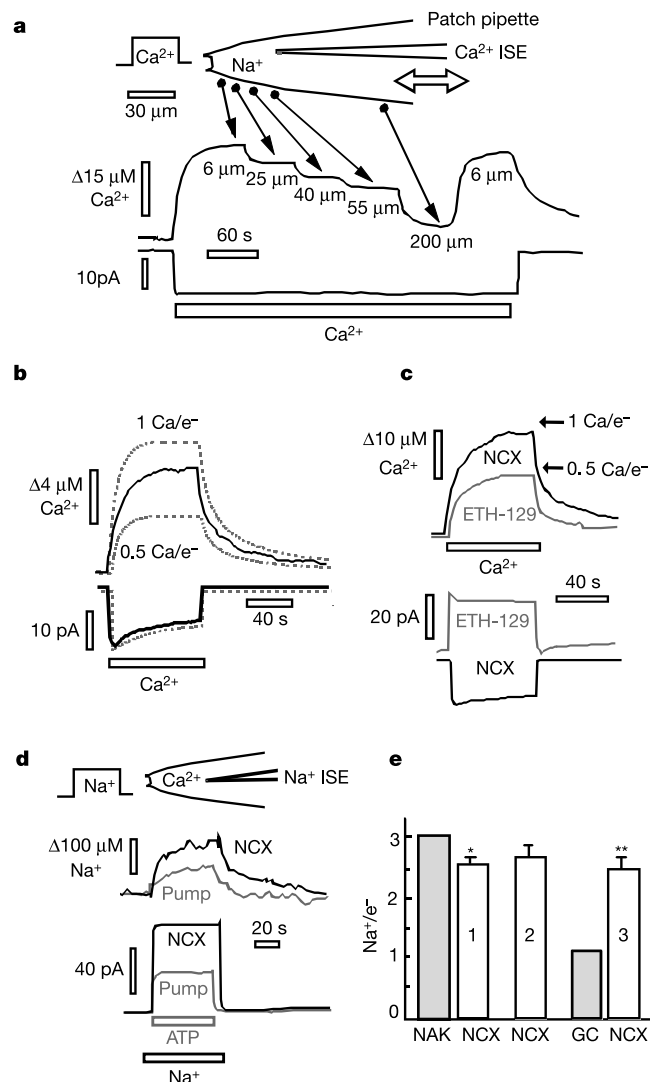


Figure 1 NCX1 ion-flux coupling probed with ISEs within the giant patch pipette. ISE responses are converted to concentration change. **a**, Ca_o gradient induced by inward exchange current in a BHK cell patch. The Ca^{2+} ISE tip was manipulated from the front of the patch electrode to the four indicated positions, and again to the front before removing Ca_i . **b–e**, Ca_o and Na_o gradients induced by NCX1 using guinea-pig myocyte patches. **b**, Measured and predicted extracellular Ca_o transients (upper records) and NCX1 current (lower record). Simulated current is given as a dotted line, whereas measured current is a solid line. **c**, NCX1- and ETH-129-mediated currents and Ca^{2+} transients. Arrows indicate Ca_o responses for NCX1 expected from the ETH-129 result. **d**, Na_o transients (upper records) during Na^+/K^+ pump and NCX1 activity in the same patch. **e**, NCX1 Na^+ flux coupling determined by comparison to Na^+/K^+ pump (NAK) current (1), by simulation of Na_o diffusion (2), and by comparison to gramicidin (GC)-mediated Na^+ current (3). Asterisk, $P < 0.05$; double asterisk, $P < 0.01$.

and $62\ \mu\text{m}$, respectively. Then, Ca^{2+} flux is calculated for neighbouring measurements from the mean cross-sectional areas ($A = \pi d_i d_{i+1}/4$), the Ca^{2+} gradients ($\Delta\text{Ca}_o/\Delta x$) and the temperature-adjusted Ca^{2+} diffusion coefficient¹⁶ (D) of $0.81 \times 10^{-5}\ \text{cm}^2\ \text{s}^{-1}$. From the Ca^{2+} fluxes ($AD\Delta\text{Ca}_o/\Delta x$) and current magnitude, we calculate how many Ca^{2+} ions move for each net charge moved across the membrane (Ca^{2+}/e^-). A 3:1 $\text{Na}^+/\text{Ca}^{2+}$ ratio predicts $1\ \text{Ca}^{2+}/e^-$, whereas a 4:1 ratio predicts $0.5\ \text{Ca}^{2+}/e^-$. We found that the average was 0.91, and it was 0.88 ± 0.11 for five experiments. Assuming that NCX1 transports only Ca^{2+} and Na^+ , the $\text{Na}^+/\text{Ca}^{2+}$ transport ratio (n) can be calculated from the equation $\text{Ca}^{2+}/e^- = 1/(n - 2)$. Thus, the average transport ratio is 3.14.

Figure 1b–d shows results with cardiac patches for both forward and reverse Ca^{2+} transport. Patches were pretreated with chymotrypsin ($1\ \text{mg}\ \text{ml}^{-1}$ for 1 min) to destroy regulatory reactions that inhibit reverse transport in the absence of Ca_i (ref. 17). Figure 1b shows the time course of ΔCa_o when forward transport is activated and turned off, together with simulations of ΔCa_o . The ISE was placed $40\ \mu\text{m}$ from the pipette orifice ($d = 15\ \mu\text{m}$), and recordings are converted from millivolts to Ca_o . Current reaches a peak at $-15\ \text{pA}$ and decays by 40% over 2 min while ΔCa_o reaches $9.4\ \mu\text{M}$. For the simulations¹⁰, current was reconstructed as an exponential decay to a plateau (Fig. 1b, dotted line), pipette dimensions were

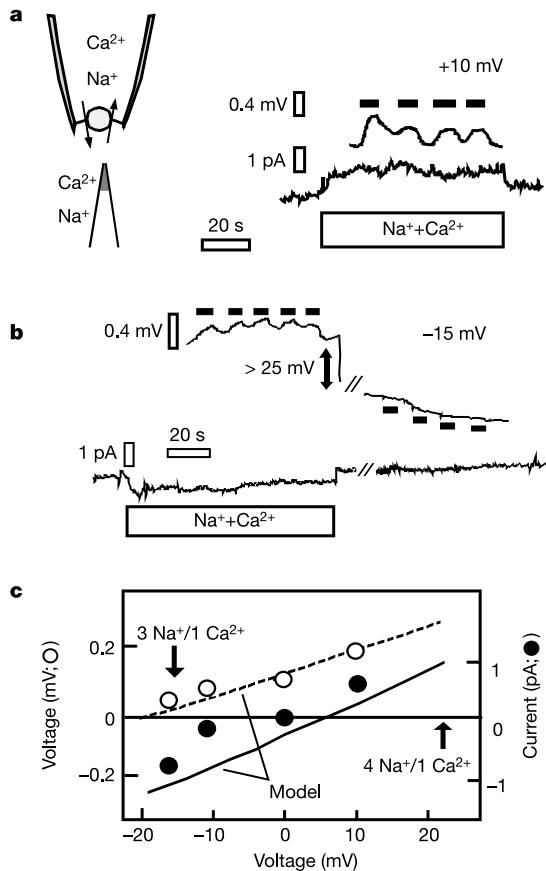


Figure 2 Reversal potentials of Ca^{2+} flux and current in a giant cardiac membrane patch. The patch pipette ($120\ \text{mM}\ \text{Na}_o$, $1.3\ \text{mM}\ \text{Ca}_o$) is moved laterally in front of a Ca^{2+} ISE to detect Ca_i gradients induced by applying $12\ \text{mM}\ \text{Na}_i$ and $0.8\ \mu\text{M}$ free Ca_i ($20\ \mu\text{M}$ EGTA). Na^+ was replaced with Cs^+ . **a**, Exchange current is outward at $+10\ \text{mV}$ (**a**) and inward at $-15\ \text{mV}$ (**b**). A higher Ca_i concentration is detected close to the pipette orifice at both potentials. **c**, Composite results with the voltage dependence of Ca^{2+} flux (dotted line) and charge flux (solid line) predicted by the NCX1 model described in Fig. 3. At $+20\ \text{mV}$, the simulated Ca^{2+} transport rate is $380\ \text{s}^{-1}$, and the simulated charge transfer rate is $180\ \text{s}^{-1}$ with the ion concentrations used in these experiments.

reconstructed over $2\ \text{mm}$, and the patch was assumed to be $5\ \mu\text{m}$ from the orifice, as observed in the microscope. Predicted Ca_o transients, given for 1 and $0.5\ \text{Ca}^{2+}/e^-$, recreate very well the time course of Ca_o changes. Thus, D is accurate. A Ca^{2+}/e^- value of 0.80 gave the best fit, which gives a transport ratio of 3.25 . The average ratio was 3.3 ± 0.2 ($n = 8$).

In a third approach, we compared Ca^{2+} transport by NCX1 and an electrogenic Ca^{2+} ionophore, ETH-129 (ref. 18), in the same patches. The ionophore is expected to transport $0.5\ \text{Ca}^{2+}/e^-$, and we measured an average value of 0.47 ($n = 4$) in Na^+ -free solutions. In Fig. 1c, NCX1 activity was recorded first ($0.1\ \text{mM}\ \text{Ca}_i$). Then, Ca^{2+} transport by ETH-129 ($30\ \mu\text{M}$) was recorded with $140\ \text{mM}$

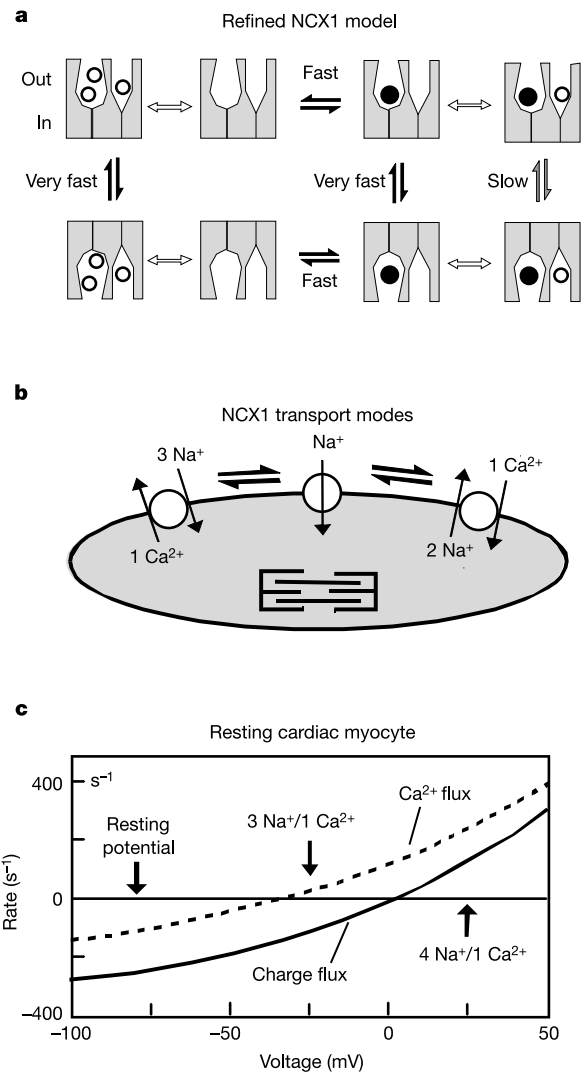


Figure 3 Refined $\text{Na}^+/\text{Ca}^{2+}$ exchange model. **a**, State diagram. Each NCX1 cartoon is oriented with the extracellular side up. $2\ \text{Na}^+$ or $1\ \text{Ca}^{2+}$ bind at one site, whereas only $1\ \text{Na}^+$ binds at the second site. When binding sites are occupied by $3\ \text{Na}^+$ or $1\ \text{Ca}^{2+}$, ion translocation is very fast ($30,000\ \text{s}^{-1}$). Translocation of $1\ \text{Na}^+$ together with $1\ \text{Ca}^{2+}$ occurs approximately ten times slower ($2,500\ \text{s}^{-1}$), and Ca^{2+} release rates are intermediate ($10,000\ \text{s}^{-1}$). Three exchange modes are possible. Filled circle, Ca^{2+} ; open circle, Na^+ . **b**, NCX1 function in a resting myocyte. Ca^{2+} import by the $2\ \text{Na}^+/1\ \text{Ca}^{2+}$ mode is balanced by Ca^{2+} export by the $3\ \text{Na}^+/1\ \text{Ca}^{2+}$ mode. Simultaneously, the Na^+ -conducting mode generates background inward current. **c**, Predicted voltage dependence of single-exchanger Ca^{2+} flux and charge flux (that is, $I-V$) in the resting state of a cardiac myocyte. Positive y -axis values give inward Ca^{2+} flux and outward cation flux rates. Relevant to cardiac physiology, the $\text{Na}^+/\text{Ca}^{2+}$ transport ratio is about 4 in the negative potential range when free Ca_i rises to $2\ \mu\text{M}$ (not shown).

Na_i and $50 \mu\text{M Ca}_i$, conditions in which forward Ca^{2+} transport by NCX1 is inhibited by >95%. For ETH-129, the ΔCa_o -to-current ratio ($\mu\text{M/pA}$) is 0.37, and for NCX1 the ratio is 0.67. Thus, NCX1 is calculated to move $0.90 \text{ Ca}^{2+}/\text{e}^-$, which gives a $\text{Na}^+/\text{Ca}^{2+}$ transport ratio of 3.11. We conclude that the $\text{Na}^+/\text{Ca}^{2+}$ transport ratio for fully activated forward transport is about 3.2 and does not exceed 3.4.

Na^+ transport can also be monitored by means of Na^+ ISEs in the patch pipette (see Fig. 1d, e). Na_o accumulations (ΔNa_o) of >20 μM are accurately detected with 2 mM background Na_o , and we devised protocols to compare reverse NCX1 and ATP-driven Na^+/K^+ pump activities in the same patches. To do so, Na^+/K^+ pumps were activated first by applying ATP_i (1 mM) in the presence of K_o (1 mM) and Na_i (10 mM), but with no Ca_i (10 mM EGTA). Although Ca_o (1 mM) is present, no NCX1 current is activated because reverse Ca^{2+} transport requires Ca_i for activation¹⁷. Thus, pump current is monitored in isolation. Then, the patches were treated with chymotrypsin to destroy NCX1 regulation, and outward NCX1 current was recorded with 60 mM Na_i in the absence of ATP_i . The ΔNa_o -to-current ratio is 21% greater for the pump than for NCX1, and the average ratio was 12% greater for the pump ($P < 0.05$; $n = 3$). Na^+/K^+ pumps operate with a transport ratio close to 3:2 (Na^+/K^+)¹⁹; that is, 1 net charge per cycle, $3 \text{ Na}^+/\text{e}^-$. Accordingly, NCX1 transports about $2.64 \text{ Na}^+/\text{e}^-$ (number 1 in Fig. 1e). By comparison, an average ratio of $2.7 \text{ Na}^+/\text{e}^-$ (number 2 in Fig. 1e; $n = 3$) was obtained by simulating Na_o diffusion ($D = 1.38 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$; ref. 16). Finally, we compared NCX1 to the ionophore gramicidin¹⁰. As gramicidin operates as a monovalent cation channel, Na^+/e^- ratios can be forced to 1 by using solutions with Na^+ as the only permeant cation. NCX1 activity was monitored first, and thereafter the current and ΔNa_o induced by application of gramicidin (0.2 mg ml^{-1}) were determined in the presence of low Na_i (15 mM). The ΔNa_o -to-current ratio was 2.46-fold higher (± 0.11) for NCX1 than for gramicidin (number 3

in Fig. 1e; $n = 3$). The $\text{Na}^+/\text{Ca}^{2+}$ transport ratios (n), calculated from the relationship $\text{Na}^+/\text{e}^- = n/(n-2)$, are 3.22, 3.18 and 3.37 for the three types of experiment, respectively. We conclude that the $\text{Na}^+/\text{Ca}^{2+}$ transport ratio for fully activated reverse transport is close to 3.2, and cannot be greater than 3.4.

To explain why different $\text{Na}^+/\text{Ca}^{2+}$ transport ratios were inferred from Ca^{2+} flux⁶ and current reversal potentials^{7,8}, it must be known whether Ca^{2+} flux and current reverse at the same membrane potential. To determine this, Ca^{2+} ISEs were used on the cytoplasmic side to detect the direction of Ca^{2+} flux under near-equilibrium conditions. An EGTA concentration of 20 μM was carefully chosen for the cytoplasmic side to maintain an accurate free Ca_i while not preventing detection of small Ca_i gradients. In Fig. 2, with 120 mM Na_o and 1.3 mM Ca_o , NCX1 currents were activated by switching from Na_i - and Ca_i -free solution to solution with 12 mM Na_i and 0.8 $\mu\text{M Ca}_i$. Bars in Fig. 2a (+10 mV) and Fig. 2b (-15 mV) indicate times when the patch pipette orifice was moved to a point just in front of the Ca^{2+} ISE from a position 60 μm distant. Positive ISE deflections in the presence of current indicate Ca^{2+} flux to the cytoplasmic side. The lack of ISE responses in the absence of Na_i and Ca_i (Fig. 2b) indicates that the Ca^{2+} gradients detected are generated by $\text{Na}^+/\text{Ca}^{2+}$ exchange. As evident in the composite results (Fig. 2c), NCX1 current clearly reverses between +10 and -15 mV, mid-way between reversal potentials for $\text{Na}^+/\text{Ca}^{2+}$ transport ratios of 3:1 and 4:1. The Ca_i gradient from the pipette tip to the bath becomes smaller at more negative potentials, but it definitively remains positive at all potentials. This indicates that NCX1 cannot generate larger Ca^{2+} gradients than expected for a $\text{Na}^+/\text{Ca}^{2+}$ transport ratio of 3:1, and that Na^+ transported in excess of a 3:1 ratio moves as Ca^{2+} -dependent Na^+ 'slippage'.

As Na^+ and Ca^{2+} can be bound at the same time by NCX1 (refs 20, 21), the simplest possible slippage mechanism is that Na^+ and Ca^{2+} are transported together with a lower probability than 1

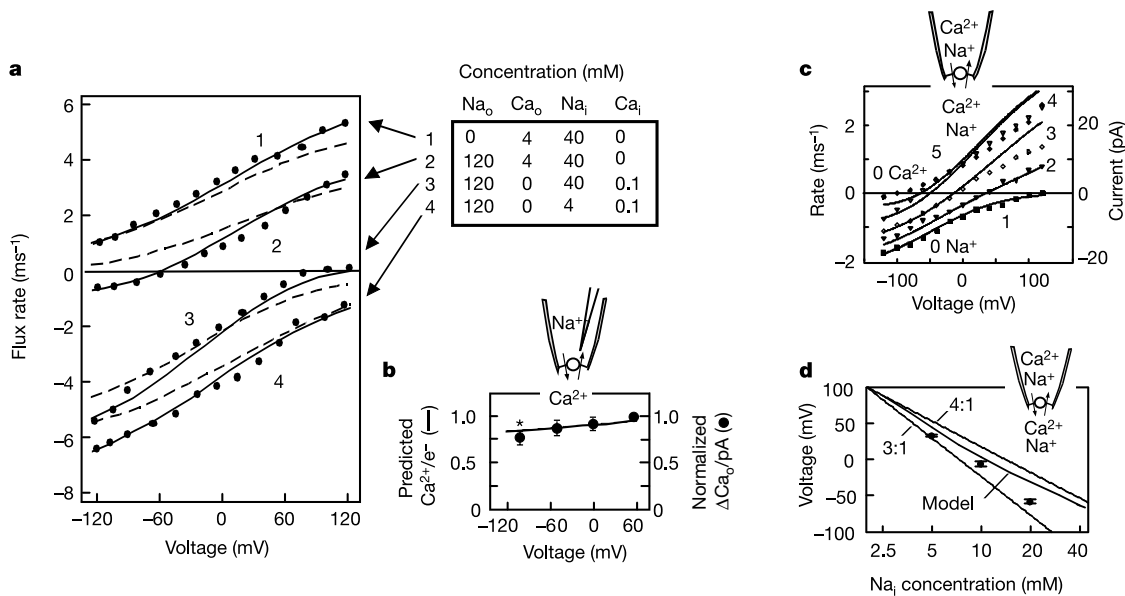


Figure 4 Voltage dependence of NCX1 flux coupling. **a**, NCX1 I - V s (dots) from BHK patches with model predictions for charge flux (solid lines) and Ca^{2+} flux (dashed lines). Positive y -axis values give outward cation and inward Ca^{2+} flux. Peak currents under the four conditions given in the right panel are +55, +20, -37 and -42 pA, respectively. Results 3 and 4 are from the same patch. **b**, Voltage dependence of the predicted $\text{Ca}^{2+}/\text{e}^-$ ratio and the measured ΔCa_0 /current ratio for Ca^{2+} export (120 mM Na_o , 0.1 mM Ca_i). Asterisk indicates $P < 0.02$, comparing values at -120 and +60 mV. The solid line is the predicted $\text{Ca}^{2+}/\text{e}^-$ value. Data points are ratios of Ca^{2+} accumulation to

current, normalized to the value at +50 mV ($n = 5$). **c**, NCX1 reversal potentials with 100 mM Na_o , 1.2 mM Ca_o and 0.5 $\mu\text{M Ca}_i$. The results are with no Na_i (1), with 5 mM Na_i (2), with 10 mM Na_i (3) and with 20 mM Na_i (4). Result 5 is with 20 mM Na_i and no Ca_i . All results are subtractions of currents with and without the indicated Na_i and Ca_i concentrations, together. Voltage was varied bi-directionally in steps and results were averaged²⁷. Solid lines are model predictions. **d**, Average reversal potentials under conditions in **c** ($n = 12$), together with predicted results for $3 \text{ Na}^+/\text{e}^-$, $4 \text{ Na}^+/\text{e}^-$ and the model.

Ca^{2+} or 3 Na^+ (ref. 22). In fact, the lines in Fig. 2c are predictions of this hypothesis²² using a model with transport rates and voltage dependencies determined previously^{14,15,21,23,24}. As shown in Fig. 3, 3 Na^+ can exchange for 1 Ca^{2+} , 1 Na^+ and 1 Ca^{2+} can exchange for 1 Ca^{2+} (that is, a Na^+ -conducting mode), and 1 Ca^{2+} and 1 Na^+ can exchange for 3 Na^+ ions (that is, a net $2 \text{ Na}^+/1 \text{ Ca}^{2+}$ exchange). In a resting myocyte (Fig. 3b), the Na^+ -conducting mode generates Na^+ import that would reverse at the Nernst potential for Na^+ . The $2 \text{ Na}^+/1 \text{ Ca}^{2+}$ mode generates only electroneutral Ca^{2+} import, as this stoichiometry does not allow export across a 10,000-fold Ca^{2+} gradient. Furthermore, the $3 \text{ Na}^+/1 \text{ Ca}^{2+}$ mode extrudes Ca^{2+} to compensate for that import. Figure 3c shows the voltage dependence of Ca^{2+} and charge flux (that is, NCX1 current) for a single exchanger in a resting myocyte (1.2 mM Ca_o , 140 mM Na_o , 0.1 μM Ca_i and 8 mM Na_i). Ca^{2+} flux reverses 8 mV more negatively than expected for a $3 \text{ Na}^+/1 \text{ Ca}^{2+}$ exchanger, whereas NCX1 current (that is, charge flux) reverses 27 mV more positively, as in cardiac macro-patches⁷.

The model described in Fig. 3 makes striking predictions that we have verified and describe in Fig. 4 using BHK cell patches. Figure 4a shows current–voltage relations (I -Vs) for four key conditions (dots), scaled to predicted single-exchanger Ca^{2+} flux-voltage (dashed lines) and charge flux-voltage (solid lines) relations (that is, I -Vs). The I -Vs for fully activated outward and inward NCX1 currents (numbers 1 and 4) are accurately reproduced, and the predicted Ca^{2+}/e^- ratio at 0 mV is about 0.9 for both. A simple NCX1 model would generate only outward current in the absence of Ca_i . However, with 4 mM Ca_o , 120 mM Na_o , 40 mM Na_i and no added Ca_i with 40 mM EGTA_i (number 2), inward current is generated by the Na^+ -conducting mode. The inward current is activated by Na_i because without Na_i (or Ca_i) all binding sites become locked open to the cytoplasmic side and cannot engage in ion transport (see Fig. 3a). Very similar I -Vs have been defined by addition and removal of Na_i , as in Fig. 4a ($n > 20$), or by application and removal of the NCX1 inhibitor KB-R7943 (L. Hrysko, personal communication).

A closely related prediction (Fig. 4a, number 3) occurs with high Na_i (40 mM) and Ca_i (0.1 mM), together with high Na_o (120 mM) and no Ca_o (40 mM EGTA). At +120 mV, the modelled current nearly reaches the point of reversal. Outward current can occur at positive potentials without Ca_o because 1 Na^+ plus 1 Ca^{2+} can be transported to the outside, followed by return of Ca^{2+} to the inside without Na^+ . For this condition, current reversal clearly occurred with multiple solution changes in 50% of patches, and reversal did not clearly occur in about 25% of patches (see ref. 21).

For forward exchange current (Fig. 4a, number 4), the predicted Ca^{2+}/e^- ratio decreased from 0.92 at +50 mV to 0.78 at −100 mV (solid line in Fig. 4b). This is because Ca^{2+} release to the extracellular side becomes limiting for Ca^{2+} transport, whereas retrograde transport of 1 Ca^{2+} with 1 Na^+ is favoured by negative potentials. This prediction was verified by analysing ΔCa_o as $\mu\text{M}/\text{pA}$ ratios at multiple membrane potentials in five highly stable patches. The ratio decreases by 21% at −100 mV compared with +50 mV ($P < 0.02$), with no difference between control ($n = 3$) and chymotrypsin-treated patches ($n = 3$).

Figure 4c illustrates I -Vs with physiologically relevant ion gradients (100 mM Na_o , 1.2 mM Ca_o , 0.5 μM Ca_i and various concentrations of Na_i). NCX1 current with 0.5 μM Ca_i and no Na_i (Fig. 4c, number 1) does not reverse, whereas reversal occurs with Na_i and no Ca_i (number 5), as described above. With 5, 10 and 20 mM Na_i and 0.5 μM Ca_i (numbers 2, 3 and 4), currents reverse at 38, −15 and −55 mV, respectively. As summarized in Fig. 4d, fourfold changes of Na_i result in an average 93 mV shift of reversal potential ($n = 12$), substantially greater than expected for $4 \text{ Na}^+/1 \text{ Ca}^{2+}$ coupling (74 mV) and substantially smaller than expected for $3 \text{ Na}^+/1 \text{ Ca}^{2+}$ coupling (110 mV).

Finally, we point out that ion-conducting modes of NCX1, if they exist, may also explain our findings. But for now, $\text{Na}^+/\text{Ca}^{2+}$ co-transport explains all discrepancies from a pure $3 \text{ Na}^+/1 \text{ Ca}^{2+}$ transport function in a conservative and specific fashion. Co-transport of Ca^{2+} with K^+ by neuronal $\text{Na}^+/\text{Ca}^{2+}$ exchangers²⁵ forms a precedent for our hypothesis. Differences between experimental results in patches^{7,21}, intact myocytes²⁶ and BHK versus HEK293 (ref. 8) cells suggest that cellular environment modulates NCX1 slippage. $\text{Na}^+/\text{Ca}^{2+}$ co-transport by NCX1 may determine why resting myocyte Ca_i is higher than expected for a $3 \text{ Na}^+/1 \text{ Ca}^{2+}$ ratio²², and it may contribute background inward current that determines resting Na_i and supports cardiac pace-making. □

Methods

Cells, electrophysiology and ion-selective electrodes

Guinea-pig myocyte preparation and storage²⁷; maintenance of BHK cells expressing NCX1 (ref. 13); giant patch formation from myocytes (4–8 pF)²⁷ and BHK cells (3–6 pF)²⁸; determination of I -Vs²⁷, ISE fabrication and recording¹⁰; and simulations^{10,29} were all carried out as described previously. All recordings were made at 36 °C and 0 mV unless indicated otherwise.

Solutions

For myocyte patches, Cl^- -containing solutions were used because Cl^- conductance is not detected under the conditions that we used. For BHK patches, low Cl^- (1–6 mM total) solutions were used because small background Cl^- conductances are observed. No Na^+ conductance was detected in either membrane when Cs^+ was substituted for Na^+ in solutions with at least 10 mM Cs^+ , 10 mM TEA and 1 mM EGTA. Li^+ was not used as a substitute for Na^+ , because Li^+ tended to destabilize seals over time and/or induce a small conductance.

The pipette solution in Fig. 1a contained (in mM): 100 NaOH, 20 HEPES, 1 MgCl_2 , 10 CsOH, 30 TEA-OH, 0.008 free Ca^{2+} calibrated with Ca^{2+} ISEs against buffered standards, pH 7.0, with aspartate. The bath solutions in Fig. 1a contained (in mM): 120 CsOH, 20 TEA-OH, 20 HEPES, 1 EGTA with or without 1.1 Ca^{2+} , pH 7.0, with aspartate. The pipette solution in Fig. 1b, c contained (in mM): 140 NaCl, 5 HEPES, 0.12 ouabain, 1 MgCl_2 , 2 CsCl, 10 TEA-Cl, 0.013 free Ca^{2+} , pH 7.0, with *N*-methylglucamine (NMG). The bath solution in Fig. 1b, c contained (in mM): 140 CsCl (or 140 NaCl with ETH-129), 10 TEA-Cl, 10 HEPES, 1 EGTA with or without 1.1 Ca^{2+} , pH 7.0, with NMG. The pipette solution in Fig. 1d, e contained (in mM): 120 NMG, 20 HEPES, 1 CaCl_2 , 2 NaCl, 0.1 ouabain or 1 KCl (when Na^+/K^+ pump was recorded), pH 7.0, with HCl. For NCX1 current, the bath solution in Fig. 1d, e contained (in mM): 20 CsCl, 40 TEA-OH, 10 EGTA, 1 EDTA, 1.5 MgCl_2 , 20 HEPES and either 60 NaCl or 37 TEA-Cl plus 23 CsCl, pH 7.0, with HCl. For the Na^+/K^+ pump, bath solution in Fig. 1d, e contained (in mM): 20 CsCl, 40 TEA-OH, 10 EGTA, 1 EDTA, 1.5 MgCl_2 , 20 HEPES, 50 NMG, 10 NaCl with or without 1 Mg-ATP , pH 7.0, with HCl. The pipette solution in Fig. 2 contained (in mM): 120 NaCl, 1.3 CaCl_2 , 10 CsCl, 10 TEA-Cl and 10 HEPES, set to pH 7.0, with NMG. The bath solution in Fig. 2 contained (in mM): 100 CsCl, 20 TEA-Cl, 0.020 EGTA with or without CaCl_2 to match 0.8 μM buffered Ca^{2+} standard, 12 NaCl or 12 additional CsCl, pH 7.0, with NMG. All solutions in Fig. 4 contained (in mM): 20 (Fig. 4a, b) or 40 (Fig. 4c, d) TEA-OH, 20 CsOH, 0.5 MgCl_2 and 20 HEPES. NaOH and CsOH were used as substitutes to achieve 150 mM total monovalent cations with Na^+ concentrations as indicated in the text and figure legend. Cytoplasmic solutions contained 10 mM EGTA with Ca-sulphamate to achieve desired free Ca^{2+} concentrations. For Ca^{2+} -free pipette and bath solution (Fig. 4a, records 2 and 3), 40 mM EGTA was used with 2 mM MgCl_2 , pH 7.0, with aspartate.

Simulations

The model described in Fig. 3a was represented as a pseudo 4-state transport model²⁹ for simulation, and all results presented use a single set of parameters. Details are available as Supplementary Information.

Received 5 September; accepted 5 December 2003; doi:10.1038/nature02271.

- Reuter, H. & Seitz, N. The dependence of calcium efflux from cardiac muscle on temperature and external ion composition. *J. Physiol. (Lond.)* **195**, 451–470 (1968).
- Philipson, K. D. & Nicoll, D. A. Sodium-calcium exchange: a molecular perspective. *Annu. Rev. Physiol.* **62**, 111–133 (2000).
- Kimura, J., Miyamae, S. & Noma, A. Identification of sodium-calcium exchange current in single ventricular cells of guinea-pig. *J. Physiol. (Lond.)* **384**, 199–222 (1987).
- Noble, D. The surprising heart: a review of recent progress in cardiac electrophysiology. *J. Physiol. (Lond.)* **353**, 1–50 (1984).
- Lytton, J., Schmetkamp, P. P. M., Hryshko, L. V. & Blaustein, M. P. *Proc. 4th Int. Conf. (New York Academy of Science, New York, 2002)*.
- Reeves, J. P. & Hale, C. C. The stoichiometry of the cardiac sodium-calcium exchange system. *J. Biol. Chem.* **259**, 7733–7739 (1984).
- Fujioka, Y., Komeda, M. & Matsuoka, S. Stoichiometry of $\text{Na}^+-\text{Ca}^{2+}$ exchange in inside-out patches excised from guinea-pig ventricular myocytes. *J. Physiol. (Lond.)* **523**, 339–351 (2000).
- Dong, H., Dunn, J. & Lytton, J. Stoichiometry of the cardiac $\text{Na}^+/\text{Ca}^{2+}$ exchanger NCX1.1 measured in transfected HEK cells. *Biophys. J.* **82**, 1943–1952 (2002).
- Smith, P. J. Non-invasive ion probes—tools for measuring transmembrane ion flux. *Nature* **378**, 645–646 (1995).

10. Kang, T. M., Markin, V. S. & Hilgemann, D. W. Ion fluxes in giant excised cardiac membrane patches detected and quantified with ion-selective microelectrodes. *J. Gen. Physiol.* **121**, 325–348 (2003).
11. DeFelice, L. J. & Galli, A. Fluctuation analysis of norepinephrine and serotonin transporter currents. *Methods Enzymol.* **296**, 578–593 (1998).
12. Mackenzie, B., Loo, D. D. & Wright, E. M. Relationships between Na^+ /glucose cotransporter (SGLT1) currents and fluxes. *J. Membr. Biol.* **162**, 101–106 (1998).
13. Linck, B. *et al.* Functional comparison of the three isoforms of the Na^+ / Ca^{2+} exchanger (NCX1, NCX2, NCX3). *Am. J. Physiol.* **274**, C415–C423 (1998).
14. Hilgemann, D. W., Nicoll, D. A. & Philipson, K. D. Charge movement during Na^+ translocation by native and cloned cardiac Na^+ / Ca^{2+} exchanger. *Nature* **352**, 715–718 (1991).
15. Hilgemann, D. W. Unitary cardiac Na^+ , Ca^{2+} exchange current magnitudes determined from channel-like noise and charge movements of ion transport. *Biophys. J.* **71**, 759–768 (1996).
16. Robinson, R. A. & Stokes, R. H. *Electrolyte solutions* (Butterworths, London, 2002).
17. Hilgemann, D. W. Regulation and deregulation of cardiac Na^+ - Ca^{2+} exchange in giant excised sarcolemmal membrane patches. *Nature* **344**, 242–245 (1990).
18. Wang, E. *et al.* Transport properties of the calcium ionophore ETH-129. *Biophys. J.* **81**, 3275–3284 (2001).
19. Rakowski, R. F., Gadsby, D. C. & De Weer, P. Stoichiometry and voltage dependence of the sodium pump in voltage-clamped, internally dialyzed squid giant axon. *J. Gen. Physiol.* **93**, 903–941 (1989).
20. Reeves, J. P. & Sutko, J. L. Competitive interactions of sodium and calcium with the sodium-calcium exchange system of cardiac sarcolemmal vesicles. *J. Biol. Chem.* **258**, 3178–3182 (1983).
21. Matsuoka, S. & Hilgemann, D. W. Steady-state and dynamic properties of cardiac sodium-calcium exchange. Ion and voltage dependencies of the transport cycle. *J. Gen. Physiol.* **100**, 963–1001 (1992).
22. Kang, T. M., Steciuk, M. & Hilgemann, D. W. Sodium-calcium exchange stoichiometry: is the noose tightening? *Ann. NY Acad. Sci.* **976**, 142–151 (2002).
23. Niggli, E. & Lederer, W. J. Molecular operations of the sodium-calcium exchanger revealed by conformation currents. *Nature* **349**, 621–624 (1991).
24. Kappel, M. & Hartung, K. Rapid charge translocation by the cardiac Na^+ - Ca^{2+} exchanger after a Ca^{2+} concentration jump. *Biophys. J.* **71**, 2473–2485 (1996).
25. Cervetto, L., Lagnado, L., Perry, R. J., Robinson, D. W. & McNaughton, P. A. Extrusion of calcium from rod outer segments is driven by both sodium and potassium gradients. *Nature* **337**, 740–743 (1989).
26. Hinata, M. *et al.* Stoichiometry of Na^+ - Ca^{2+} exchange is 3:1 in guinea-pig ventricular myocytes. *J. Physiol. (Lond.)* **545**, 453–461 (2002).
27. Hilgemann, D. W., Matsuoka, S., Nagel, G. A. & Collins, A. Steady-state and dynamic properties of cardiac sodium-calcium exchange. Sodium-dependent inactivation. *J. Gen. Physiol.* **100**, 905–932 (1992).
28. Hilgemann, D. W. & Lu, C. C. Giant membrane patches: improvements and applications. *Methods Enzymol.* **293**, 267–280 (1998).
29. Hilgemann, D. W. Numerical approximations of sodium-calcium exchange. *Prog. Biophys. Mol. Biol.* **51**, 1–45 (1988).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Feng and P. Foley for technical assistance; K. Philipson for providing the BHK cell line; L. Hrysko for defining NCX1 I - V s with organic NCX1 inhibitors; and A. Ferguson and L. DeFelice for critical comments on the manuscript. This work was supported by an NIH grant to D.W.H. and a Samsung Biomedical Research Institute grant to T.M.K.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.W.H. (donald.hilgemann@ucla.edu).

A proton pore in a potassium channel voltage sensor reveals a focused electric field

Dorine M. Starace & Francisco Bezanilla

Department of Physiology and Department of Anesthesiology, David Geffen School of Medicine at UCLA, Los Angeles, California 90095, USA

Voltage-dependent potassium channels are essential for the generation of nerve impulses¹. Voltage sensitivity is conferred by charged residues located mainly in the fourth transmembrane segment (S4) of each of the four identical subunits that make up the channel. These charged segments relocate when the potential

difference across the membrane changes^{2,3}, controlling the ability of the pore to conduct ions. In the crystal structure of the *Aeropyrum pernix* potassium channel KvAP⁴, the S4 and part of the third (S3B) transmembrane α -helices are connected by a hairpin turn in an arrangement termed the 'voltage-sensor paddle'. This structure was proposed to move through the lipid bilayer during channel activation, transporting positive charges across a large fraction of the membrane⁵. Here we show that replacing the first S4 arginine by histidine in the *Shaker* potassium channel creates a proton pore when the cell is hyperpolarized. Formation of this pore does not support the paddle model, as protons would not have access to a lipid-buried histidine. We conclude that, at hyperpolarized potentials, water and protons from the internal and external solutions must be separated by a narrow barrier in the channel protein that focuses the electric field to a small voltage-sensitive region.

The recently solved complete crystal structure of KvAP⁴ did not clarify how the voltage sensor of an ion channel moves charge across the membrane, because the structure was incompatible with the authors' functional data⁵. To reconcile the structural and functional data for KvAP, the authors combined parts of two KvAP structures that were crystallized in different conditions, and used these data to produce their paddle model of activation⁵. This model raises important questions: whether the charged voltage-sensing residues reside in the bilayer, and how far they move when changing from the resting to the active position. These questions can be answered by titrating the voltage-sensing residues with protons by varying the pH of the external and internal medium and then measuring the resulting movement of gating charge. The pK_a values of the naturally occurring arginines are too basic for such a study, so we replaced these residues individually with histidine. Upon histidine replacement of all S4 charges except for the most extracellular one (R362), the S4 acts as a proton transporter^{6,7}. Here we extend our study to R362. We find that a histidine at this position also senses voltage and displaces protons, but in a very different way: rather than transporting protons, it forms a proton pore.

Histidine replacement of the first S4 charge, R362, was made in a non-conducting (W434F)⁸, non-inactivating (IR, $\Delta 6$ –46)⁹ *Shaker* potassium channel. Figure 1 shows pulse-evoked currents from the R362H mutant recorded using the macropatch technique. In an inward proton gradient, there is a large inward proton current, I_H , whose amplitude at each test pulse is independent of the membrane holding potential (compare Fig. 1a with Fig. 1b). At a holding potential of -90 mV, a steady inward current of 290 pA was recorded (Fig. 1a). Upon return to -90 mV, only pulses greater than -10 mV gave rise to tail currents with opening kinetics, indicating that a channel was fully open at potentials less than -10 mV. In contrast, when the holding potential was 0 mV, there was no steady current and the pulse-elicited currents were activated more quickly as the pulse became negative (Fig. 1b). The voltage dependence of the current, which was measured 40 ms after pulse onset, shows that at negative potentials the currents are essentially ohmic (Fig. 1c), behaviour that is typical of channel conduction.

The conductance of the R362H channel, $G(V)$, is equal to $I_H(V - E_H)$, where V is test pulse potential and E_H is the H^+ equilibrium potential. Conductance increases sharply at approximately -10 mV, and is most voltage dependent between -100 and -10 mV (Fig. 1d). The voltage dependence of conductance is much like a mirror image of the voltage dependence of gating-charge movement, Q , of the *Shaker* potassium channel (Q - V curve; see Fig. 5a). This suggests that the proton current is coupled to voltage-sensor movement. Moreover, the negative shift of -34 mV along the voltage axis in the G - V curve obtained from a holding potential of 0 mV relative to -90 mV (Fig. 1d) corresponds to the same holding-

potential-dependent shift characteristic of the Q - V curve¹⁰. Finally, the time course of R362H proton channel opening (Fig. 1b) is similar to the time course of the gating-charge movement recorded from the non-conducting *Shaker* channel when held at 0 mV (ref. 10).

The inward rectification of the R362H channel is not due to the H^+ equilibrium potential because even in the presence of an outward proton gradient, the voltage dependence of the proton currents shows inward rectification (Fig. 2a, circles and triangles). Proton gradients were imposed by varying the external pH (pH_o) and keeping the internal solution constant. Because control of internal pH (pH_i) is not precise in the cut-open oocyte voltage-clamp configuration, pH_i was measured with a proton-selective microelectrode. Figure 2b shows three series of R362H tail currents recorded in different proton gradients. When the R362H channel is open at the end of the hyperpolarizing pulse to -120 mV and the membrane potential is suddenly changed, the elicited, instantaneous tail currents, I_{inst} , are generated by the flow of ions through the open channel down their electrochemical gradients, along with a gating-current component. The gating component was removed from the tail currents by subtracting currents recorded in the presence of 5 mM $NiSO_4$, which blocks proton current through the R362H channel (see below). The voltage dependence of the instantaneous tail currents shows that, for each of the three proton gradients, the tail currents reverse direction (E_{rev}) around the Nernst H^+ equilibrium potential (Fig. 2b). Therefore, the ionic current through the open R362H channel is driven by the H^+ electrochemical gradient, indicating that the R362H current is carried by protons. The protons flow down their electrochemical gradient when the pore is open, so rectification must be imposed by

a voltage-dependent mechanism that closes the pore at positive potentials. The most obvious candidate is the introduced histidine: it is charged and because it replaces one of the voltage-sensing residues, it is expected to move in a voltage-dependent way. If so, then hyperpolarized potentials must drive the histidine to a position that creates a proton pore either by bridging the internal and external solutions or by connecting a proton wire. Conversely, depolarized potentials move the histidine away from this narrow transmembrane span or wire and thereby abolish or close the proton pore.

The R362H channel currents can be blocked by the histidine-binding metal nickel, indicating that the proton currents are specific to histidine. Figure 2c shows the voltage-dependent behaviour of the proton currents in pH_o 9.2, 7.4 and 5 before the addition of Ni^{2+} (filled symbols). Addition of 5 mM $NiSO_4$ to the external solution at pH_o 7.4 produced a significant reduction of the proton current (open triangles) indicating direct exposure of the histidine residue to the external solvent. The nickel block could not be removed by washing with an external solution of pH_o 7.4. The block was relieved only by washing with an external solution of pH_o 5 (open squares), indicating that an excess of protons was necessary to compete for the nickel-binding site.

To further verify the involvement of histidine in the R362H proton currents, we examined the channel pK_a by measuring current amplitude as a function of external pH. As shown in Fig. 2d, the modulation of the proton currents by pH_o at a given membrane potential (-130 mV) fitted a two-state model (Henderson-Hasselbach equation) with a pK_a of 6.5, which is consistent with the titration of a histidine residue that is considerably exposed to the external solvent.

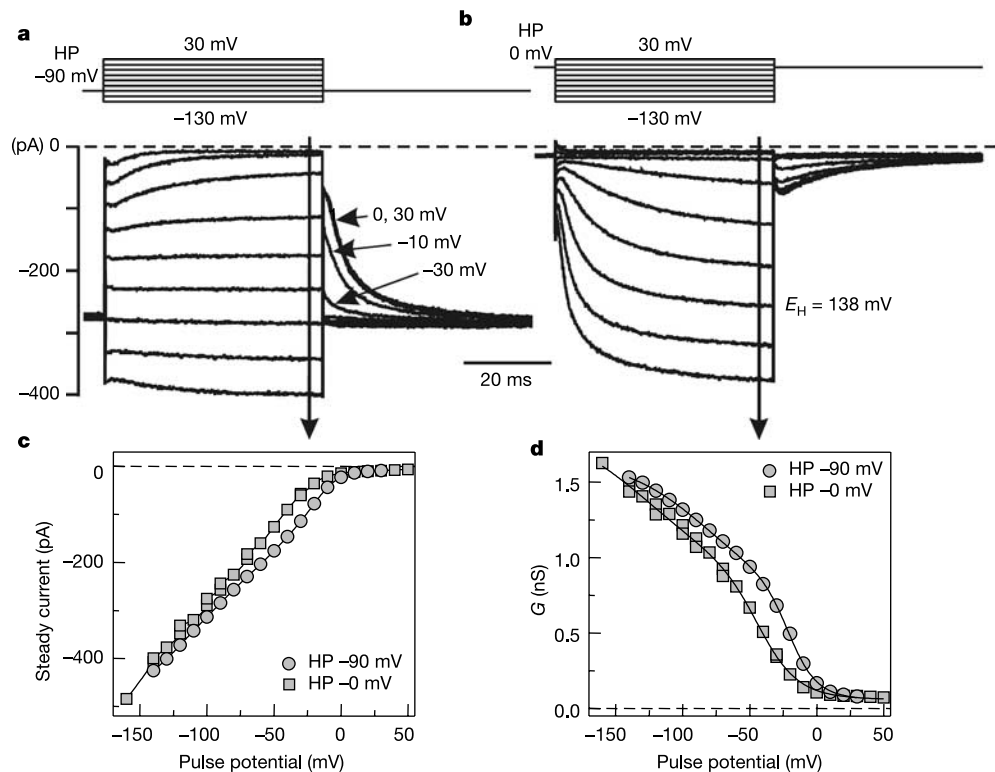


Figure 1 Voltage dependence of proton currents from the R362H *Shaker* channel. Proton currents were measured from a cell-attached macropatch on a perforated oocyte in pH_o 5/ pH_i 7.4 solutions. The superimposed currents were elicited by various test pulses, V_t , from a holding potential (HP) of **a**, -90 mV or **b**, 0 mV. **c**, **d**, Voltage dependence of the

steady-state proton current amplitudes (indicated by arrows in **a** and **b**) (I - V curves, **c**) and the proton conductance $G(V_t) = I(V_t)/(V_t - E_H)$ (G - V curves, **d**) recorded with a holding potential of -90 mV (circles) and 0 mV (squares). G - V curves were fitted to a sum of two Boltzmann functions (lines).

The voltage at which the proton current turns on increases with decreasing external pH (Fig. 2a, c), another behaviour suggesting that R362H pore formation is coupled to voltage-sensor movement. At pH_o of 7.4, the R362H proton current turns on at approximately -80 mV, a potential at which most of the gating charge is in the closed position, whereas at pH_o 5.1, the currents turn on at approximately -30 mV. This displacement of about 20 mV per pH unit is consistent with the behaviour of voltage-sensor movement, which also shifts positively along the voltage axis with decreasing pH_o owing to surface-charge screening by external protons (by about 13 mV per pH_o unit for another mutant)⁷.

The R362H channel proton currents are large, making it imperative that we prove that there is no component of the current going through the K^+ pore of the channel. We therefore examined a 'K-conducting' version of the R362H channel containing a functional K^+ pore with the mutation T449Y (rather than W434F), which increases the affinity of the K^+ pore for the blocker agitoxin II¹¹. Figure 3a shows a superimposed set of K^+ currents through the 'K-conducting' R362H channel elicited by the corresponding superimposed test-pulse protocol shown to the right of Fig. 3b, all recorded in solutions containing 120 mM K^+ inside and 2 mM K^+ outside. These currents are robust outward K^+ currents whose voltage dependence and activation kinetics are similar to those of the wild-type *Shaker* potassium channel (with fast inactivation removed). When the channel blocker agitoxin II was added to the external solution, the K^+ pore of the 'conducting' R362H channel was totally blocked (Fig. 3b). The remaining current could still be modulated by the external pH (Fig. 3c) with a pK_a of 6.2 (Fig. 3e), which is indicative of a proton current via the 362H-binding site. Moreover, in all pH gradients (Fig. 3d), the voltage dependence of

this remaining current showed the inward rectification observed in the R362H channel containing the K^+ -conduction-abolishing mutation W434F. Comparison of the proton current properties in the agitoxin-II-blocked K-conducting channels and the non-conducting R362H channels indicates that the R362H proton pore is distinct from the K^+ pore.

The behaviour of noise fluctuations in the current at steady state was examined because it reflects the underlying microscopic process and can help to distinguish between mechanisms of pore conductance and transport^{12,13}. The top of Fig. 4a shows an ensemble of 149 superimposed R362H currents in response to a -130 mV pulse from a holding potential of 0 mV. The power spectrum of the R362H proton channel, calculated from the Fourier transform of the fluctuations of the stationary current, is plotted in Fig. 4a. Simple mechanisms of conduction through a pore give rise to power spectra with lorentzian shapes^{12,13}, whereas simple mechanisms of transport generate power spectra that rise and plateau with increasing frequency, f (refs 12, 13). The R362H power spectrum fits the expected lorentzian decay of conduction through a pore¹³ with a small $1/f$ component¹⁴ (Fig. 4a, line). The rate constant of the rising phase of the proton current is 3.6 ms, which is expected to yield a lorentzian corner frequency of 44 Hz. This is very close to the corner frequency of the fitted lorentzian decay, 33 Hz (Fig. 4a, top), indicating that the opening of the proton conductance is the main gating process at the tested potential.

To estimate the conductance of a single proton channel, we performed non-stationary ensemble mean-variance analysis of R362H channel currents¹⁵. The top of Fig. 4b shows the mean and variance from a set of 142 R362H currents elicited by repetitive

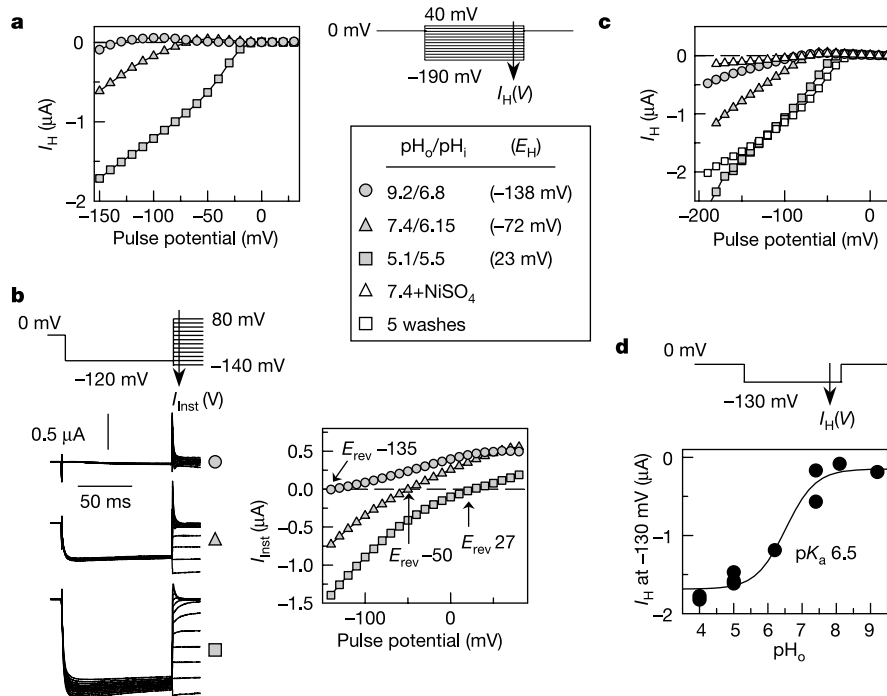


Figure 2 The R362H *Shaker* channel currents are carried by protons and are specific to histidine. I - V curves of **a**, steady-state proton currents, I_H , and **b**, instantaneous tail currents, I_{inst} , recorded in various pH gradients (indicated in box). I_{inst} - V curves (**b**, right) were obtained after subtracting currents (left) recorded in the presence of nickel to eliminate gating currents. Cut-open oocyte voltage-clamp currents were recorded from the same membrane area using the associated pulse protocol shown. The arrow in each

pulse protocol represents the isochronal point used to plot current amplitudes for each I - V curve. **c**, I - V curves of steady-state proton currents recorded, as described in **a**, before, during and after NISO₄ addition. **d**, Proton currents recorded in response to a -130 mV pulse in various values of pH_o . The I_H - pH_o curve was fitted to the Henderson-Hasselbalch equation with a pK_a of 6.5 : $I = I_{\text{min}} + (I_{\text{max}} - I_{\text{min}}) / [1 + \exp(2.3(\text{pH}_o - \text{pK}_a))]$.

–100 mV test pulses from 0 mV. Fitting the mean-variance curve shown in Fig. 4b revealed that the single-proton-channel current, i , at –100 mV was 9 fA, which yields a single-channel conductance of 40 fS (47 ± 13 fS for $n = 5$ experiments). Therefore, at –100 mV, there is a flow of protons at the rate of $i/(ze_0) \approx 56,000 \text{ s}^{-1}$ (where z is the proton valence and e_0 is the electronic charge), which is markedly larger than the value expected of a current arising from transport. In the likely case that the probability of the channel being open at –100 mV is less than 1, the flow rate would be even higher. Both the shape of the R362H channel power spectrum (Fig. 4a) and the conductance of a single proton channel are more consistent with a proton current arising from conduction through a pore rather than from transport. Moreover, the single-channel conductance of each of the four R362H proton channels in the *Shaker* tetramer is about one-quarter that of the proton channel found in eosinophil¹⁶, and is the same as the proton conductance of the gramicidin channel¹⁷.

The properties of the histidine-mediated current of R362H presented here are different from those of the voltage-sensor-coupled proton transport that arises from histidine replacement of the other *Shaker* potassium channel voltage-sensing residues, R365, R368 or R371 (refs 6, 7). An example of the voltage dependence of the proton transport current from the R365H channel is shown in Fig. 5b. The bell shape of the I_H – V curve indicates proton transport. Transport current is maximal in the voltage region where gating-charge movement, Q , is steeply voltage dependent (Fig. 5a). In this region, the histidine, coupled to the voltage sensor, makes frequent transitions between open and closed states and can transport protons, one at a time, across the membrane, with each transition in the direction of the H^+ electrochemical gradient. The transport current drops to zero at very hyperpolarized and depolarized potentials, where the histidine, coupled to the voltage sensor, favours one state and transitions to

other states become minimal. By contrast, the voltage dependence of the R362H proton currents is not bell-shaped and does not saturate at hyperpolarized potentials (Fig. 5c), a behaviour like that of an inwardly rectifying channel. We have provided evidence that the R362H currents arise from conduction through a voltage-gated pore rather than from transport with each voltage-sensor stroke. Even if conduction involves a proton-shuttle mechanism in which the histidine side chain rotates at equilibrium to alternative positions, such a conduction mechanism is distinct from voltage-driven transport.

Conduction or transport of protons by the histidine-tagged voltage sensor has important implications for its conformational changes. For proton transport to occur (Fig. 5b), the histidine must move from internal to external access with each voltage-sensor transition. Such access may occur through proton wire networks in the protein or through aqueous crevices that penetrate the hydrophobic environment and reach the voltage-sensing residues. The latter possibility is supported by other experimental evidence^{6,7,18,19}. Figure 5d shows a model of voltage-sensor movement in which the histidine-replaced residues can transport protons; depolarization of the membrane moves the histidine (filled circle), along with the other voltage-sensing residues (open circles), from an internally facing aqueous crevice to an externally facing cavity. The transfer of

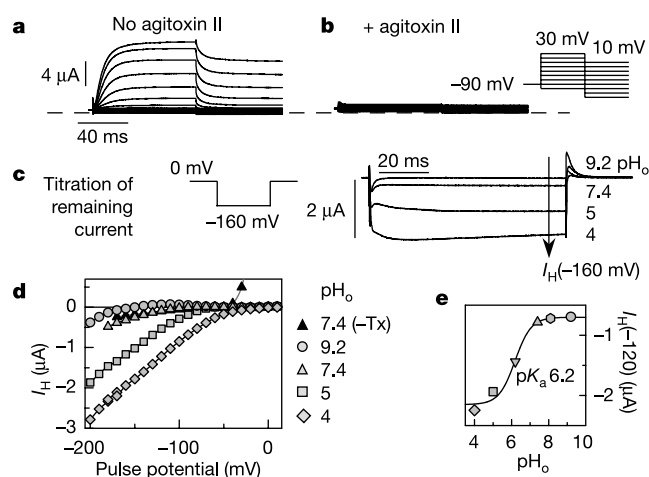


Figure 3 The R362H proton channel is distinct from the potassium channel. Cut-open voltage clamp K^+ currents recorded from the 'K-conducting' version of the R362H channel before (a) and after (b) external agitoxin II addition to a concentration of 1 μM . The pulse protocol is shown on the right. Currents were recorded in pH_0 7.4, 120 mM K^+ inside and 2 mM K^+ outside. c, pH_0 titratable currents elicited by hyperpolarizing test pulses from 0 mV remained after agitoxin block of the K^+ pore. d, I – V curves of steady-state proton currents, I_H , recorded in different pH gradients generated from pulse-elicited currents from 0 mV after agitoxin block. (Agitoxin was absent in the '–Tx' curve, from –90 mV.) e, After the agitoxin block, pH_0 modulation of proton currents elicited by –120 mV pulses fitted the Henderson–Hasselbalch equation with a pK_a of 6.2.

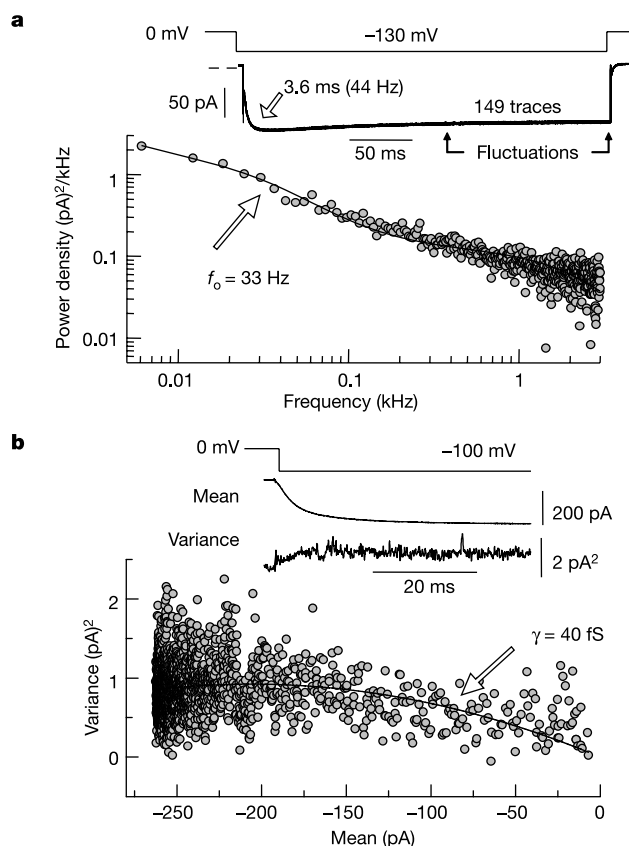


Figure 4 Biophysical properties reveal the nature of the R362H proton current. a, Top, an ensemble of 149 R362H currents recorded in pH_0 5/ pH_i 9.2, elicited by repetitive pulses to –130 mV from a holding potential of 0 mV. Bottom, power spectrum of the stationary current fluctuations. b, Top, time course of the mean and variance from a set of 142 R362H currents, recorded in pH_0 5/ pH_i 7.4 and elicited by repetitive –100 mV pulses from 0 mV. Bottom, plot of the mean versus variance at each time point, and the best fit to a parabolic function (line) predicts a single-proton-channel current of $i = 9$ fA and $N = 46,000$ proton channels.

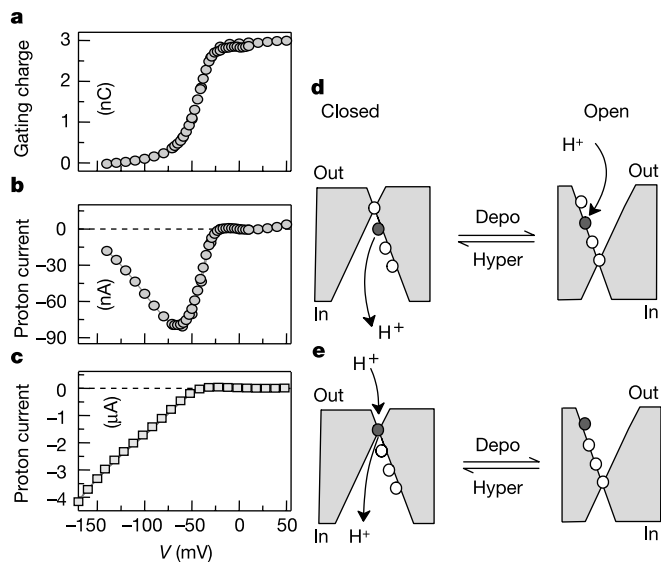


Figure 5 Proton transport and conduction: models of the conformational changes of the voltage sensor. **a**, Voltage dependence of the gating-charge movement of the *Shaker* R365H channel. **b**, Voltage dependence of proton-transport current of the R365H channel. **c**, Voltage dependence of proton channel current of the R362H channel. **d**, Conformational changes during proton transport. In a proton gradient (inward), the histidine at 365 (filled circle) transports one proton each time the voltage sensor goes from the open to the closed conformation. **e**, Representation of the R362H channel proton pore. Internal and external protons have simultaneous access to the histidine (filled circle) only in the closed position, which results in a continuous proton current when the sensor is closed. The protein core is grey. 'Depo' and 'hyper' indicate a more positive and a more negative membrane potential, respectively.

charge across the membrane in this model may be achieved by a relatively small conformational change^{20,21}.

Proton conduction by R362H imposes a particularly stringent constraint on voltage-gating models because, in the open proton pore, site 362 has to be simultaneously accessible from both sides of the membrane. This constraint is shown in Fig. 5e: when the R362H proton channel is open at hyperpolarized potentials, the histidine bridges the two water-filled crevices, thus creating a proton-selective pore. Depolarization closes the proton channel by moving 362H away from the pore-permissive bridge. The narrow membrane span necessary to create a proton pore at 362H implies that the transmembrane electric field is concentrated near the histidine. Such increased field strength has been measured directly with site-directed voltage-sensitive probes¹⁹. The focused field makes the structure exquisitely voltage sensitive and places constraints on the conformational rearrangements that the voltage sensor undergoes. A small conformational change can move the large amount of necessary gating charge across the focused transmembrane field. When histidine replaced the fourth S4 charge (R371H), both proton transport and conduction were observed⁶. However, proton conduction was small and occurred only at very depolarized potentials, a result that has also been incorporated into our model (Fig. 5d, e).

The results and implications of the R362H proton pore can be put in the context of the recently obtained crystal structures of the KvAP channel⁴. In the paddle model of channel activation, the charged residues of S4 are not exposed to the intracellular solution at all because the authors found that the S3B–S4 loop is not internally accessible to site-specific Fab fragments⁵, nor are there potential proton wire pathways in the closed state. The paddle model would require drastic modifications to accommodate movement of residues at positions 365, 368 and 371 across the bilayer from internal to

external proton access^{6,7}. One modification could be a deformation of the bilayer to allow a closer approach of the charged residues to protons that may penetrate with water at the edges of the membrane. However, even a modified paddle model could not be reconciled with the ability of the R362H channel to form a proton pore, because the integrity of the protein/bilayer in the paddle model is not broken by aqueous crevices to allow the required proton access to 362H from both sides of the membrane at negative potentials (Fig. 5e). It is important to note that the paddle model is based on docking the crystal structure of the isolated KvAP S1–S4 fragment to the pore region of the entire KvAP crystal in a particular conformation that is not unique⁵. As many other conformations are clearly possible, molecular modelling constrained by functional and structural data will most probably predict other structures that will expand the picture of possible voltage-sensor movements during activation, so that a model consistent with the constraints imposed by the R362H proton pore will emerge. One such structure, which is qualitatively compatible with our results, proposes that S4 adopts a transmembrane orientation and is packed against the pore domain near the interface between adjacent subunits²².

The results presented here on the behaviour of a proton pore created artificially from the voltage sensor of a potassium channel may have implications beyond the structure of the voltage sensor. Operation of this proton pore has revealed a unique and fundamental structural feature that may be very useful in understanding the structure–function relationship of naturally occurring voltage-dependent proton channels²³. □

Methods

Mutagenesis and expression of channels

Unless otherwise mentioned, the clone that was used as the background template for R362H mutagenesis contained the non-conducting (W434F)⁸, non-inactivating (IR, Δ6–46)⁹ *Shaker* H4 potassium channel coding sequence^{7,24}. The 'K-conducting' version of the R362H channel contained the pore mutation T449Y, but not W434F. The R362H mutation was generated by overlap extension polymerase chain reaction (PCR)²⁵ followed by mutagenic fragment replacement. All sequences generated by PCR were sequenced. To express the channel, complementary RNA was transcribed from the *NotI*-linearized DNA clone (New England Biolabs) with T7 RNA polymerase (Message Machine *in vitro* transcription kit from Ambion) and 50 nl of 0.5–0.8 μg μl⁻¹ cRNA was injected into each stage-5 *Xenopus* oocyte²⁶. Injected oocytes were maintained at 18 °C in an incubation solution of 100 mM NaCl, 2 mM KCl, 1.8 mM CaCl₂, 1 mM MgCl₂, 5 mM HEPES (pH 7.3), 0.010 mM EDTA and 0.5 mM DTT. The incubation solution was changed daily.

Electrophysiology

Channel currents were measured from oocytes 3 to 6 days after the injection of channel cRNA. Currents were recorded at 20–23 °C with the cut-open oocyte voltage-clamp technique^{6,27} or with the patch-clamp technique^{6,28}. Data were filtered at one-fifth the sampling frequency. Currents were recorded unsubtracted after analogue compensation of linear membrane capacitance at 0–50 mV. There was no subtraction or compensation of linear leak components while recording; these were subtracted off-line⁶. Unless otherwise stated, the external solutions contained 120 mM *n*-methylglucamine (NMDG), 2 mM CaCl₂, and either 20 mM CHES (2-(*N*-cyclohexylamino)ethanesulphonic acid, pH 9.2) or 20 mM HEPES (all other pHs); all solutions were brought to the appropriate pH with methanesulphonic acid (Fluka). The internal solutions were the same as the external solutions except that EGTA–NMDG replaced CaCl₂. The osmolality of all recording solutions was 240–260 mosM. All of the experiments reported here were reversible with respect to the effect of pH. Electrophysiological acquisition and analysis software were home-made.

In cut-open oocyte voltage clamps, the pH_i of the oocyte, monitored with an H⁺ electrode impaling the oocyte during current recordings, varies between pH 5.8 and 7.3 regardless of the pH of the internal solution used⁶. Therefore, for the recordings shown in Fig. 2a, b, a second intracellular electrode, filled with a liquid ion-exchanger resin selective for H⁺ (IE010 from WPI, Inc.), was used to measure pH_i⁶. The H⁺ electrode was mounted on the amplifier (Dagan) on the current electrode headstage for two-electrode clamp.

Stationary (Fig. 4a) and non-stationary noise analysis (Fig. 4b) were carried out using the patch-clamp technique in the cell-attached mode with oocytes that were ruptured to expose the internal contents to the bath solution. From a holding potential of 0 mV, a set of 100–300 proton current traces was recorded by repetitively hyperpolarizing the membrane to –100 to –160 mV every second. For stationary noise-fluctuation analysis, sets of repetitively pulsed currents were recorded in response to hyperpolarizing test pulses of 300 ms duration. Acquired data were filtered at 2.5 kHz. The power spectrum of R362H currents was determined from a fast Fourier transform of the final 150–160 ms of the 300 ms pulse-elicited currents. The power spectrum was fitted to a sum of two lorentzian

functions and a $1/f$ component. For non-stationary noise analysis, sets of repetitively pulsed currents were recorded in response to hyperpolarizing test pulses of 50 ms duration. Acquired data were filtered at 5–10 kHz. Mean current (I) and variance (σ^2) were calculated from successive pairs of current traces to minimize the effect of drift during the time taken to record the entire ensemble of currents. The single-proton-channel current (or the single-subunit current), i , and the number of proton channels (or subunits), N , was determined by fitting the $\sigma^2 - I$ plot to the equation $\sigma^2 = iI - I^2/N + B$, where B is a background-noise term.

Agitoxin II preparation

Synthetic agitoxin II was expressed and purified as a cleavable fusion protein in *Escherichia coli* according to the protocol described in ref. 29. The agitoxin expression clone was generously provided by A. Gross³⁰.

Received 22 August; accepted 5 December 2003; doi:10.1038/nature02270.

- Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond.)* **117**, 500–544 (1952).
- Seoh, S.-A., Sigg, D., Papazian, D. M. & Bezanilla, F. Voltage-sensing residues in the S2 and S4 segments of the *Shaker* K⁺ channel. *Neuron* **16**, 1159–1167 (1996).
- Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 Segment to gating charge in the *Shaker* K⁺ channel. *Neuron* **16**, 1169–1177 (1996).
- Jiang, Y. et al. X-ray structure of a voltage-dependent K⁺ channel. *Nature* **423**, 33–41 (2003).
- Jiang, Y., Ruta, V., Chen, J., Lee, A. & MacKinnon, R. The principle of gating charge movement in a voltage-dependent K⁺ channel. *Nature* **423**, 42–48 (2003).
- Starace, D. M. & Bezanilla, F. Histidine scanning mutagenesis of basic residues of the S4 segment of the *Shaker* K⁺ channel. *J. Gen. Physiol.* **117**, 469–490 (2001).
- Starace, D. M., Stefani, E. & Bezanilla, F. Voltage-dependent proton transport by the voltage sensor of the *Shaker* K⁺ channel. *Neuron* **19**, 1319–1327 (1997).
- Perozo, E., MacKinnon, R., Bezanilla, F. & Stefani, E. Gating currents from a non-conducting mutant reveal open-closed conformations in *Shaker* K⁺ channels. *Neuron* **11**, 353–358 (1993).
- Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Biophysical and molecular mechanisms of *Shaker* potassium channel inactivation. *Science* **250**, 533–538 (1990).
- Olcese, R., Latorre, R., Toro, L., Bezanilla, F. & Stefani, E. Correlation between charge movement and ionic current during slow inactivation in *Shaker* K⁺ channels. *J. Gen. Physiol.* **110**, 579–590 (1997).
- Gross, A. & MacKinnon, R. Agitoxin footprinting the *Shaker* potassium channel pore. *Neuron* **16**, 399–406 (1996).
- Lee, Y. W. *Statistical Theory of Communication* (Wiley, New York, 1960).
- Stevens, C. F. Inferences about membrane properties from electrical noise measurements. *Biophys. J.* **12**, 1028–1047 (1972).
- Conti, F., Neumcke, B., Nonner, W. & Stampfli, R. Conductance fluctuations from the inactivation process of sodium channels in myelinated nerve fibers. *J. Physiol. (Lond.)* **308**, 217–239 (1980).
- Sigworth, F. J. The variance of sodium current fluctuations at the node of Ranvier. *J. Physiol. (Lond.)* **307**, 97–129 (1980).
- Cherny, V. V., Murphy, R., Sokolov, V., Levis, R. A. & DeCoursey, T. E. Properties of single voltage-gated proton channels in human eosinophils estimated by noise analysis and by direct measurement. *J. Gen. Physiol.* **121**, 615–627 (2003).
- Schumaker, M. E., Pomès, R. & Roux, B. A. Combined molecular dynamics and diffusion model of single proton conduction through gramicidin. *Biophys. J.* **79**, 2840–2857 (2000).
- Islas, L. D. & Sigworth, F. J. Electrostatics and the gating pore of *Shaker* potassium channels. *J. Gen. Physiol.* **117**, 69–89 (2001).
- Asamoah, O. K., Wuskell, J. P., Loew, L. M. & Bezanilla, F. A fluorometric approach to local electric field measurements in a voltage-gated ion channel. *Neuron* **37**, 85–97 (2003).
- Yellen, G. The moving parts of voltage-gated ion channels. *Q. Rev. Biophys.* **31**, 239–295 (1998).
- Bezanilla, F. The voltage sensor in voltage-dependent ion channels. *Physiol. Rev.* **80**, 555–592 (2000).
- Laine, M. et al. Atomic proximity between S4 segment and pore domain in *Shaker* potassium channels. *Neuron* **39**, 467–481 (2003).
- DeCoursey, T. E. Voltage-gated proton channels and other proton transfer pathways. *Physiol. Rev.* **83**, 475–579 (2003).
- Schwarz, T. L., Tempel, B. L., Papazian, D. M., Jan, Y. & Jan, L. Y. Multiple potassium-channel components are produced by alternative splicing at the *Shaker* locus in *Drosophila*. *Nature* **331**, 137–142 (1988).
- Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59 (1989).
- Timpe, L. C. et al. Expression of functional potassium channels from *Shaker* cDNA in *Xenopus* oocytes. *Nature* **331**, 143–145 (1988).
- Stefani, E. & Bezanilla, F. The cut-open oocyte voltage clamp technique. *Methods Enzymol.* **293**, 300–318 (1998).
- Hamill, O. P., Marty, A., Neher, E., Sackmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–100 (1981).
- Park, C.-S., Hausdorff, S. F. & Miller, C. Design, synthesis, and functional expression of a gene for charybotoxin, a peptide blocker of K⁺ channels. *Proc. Natl Acad. Sci. USA* **88**, 2046–2050 (1991).
- García, M. L., García-Calvo, M., Hidalgo, P., Lee, A. & MacKinnon, R. Purification and characterization of three inhibitors of voltage-dependent K⁺ channels from *Leiurus quinquestriatus* var. *hebraeus* venom. *Biochemistry* **33**, 6834–6839 (1994).

Acknowledgements This work was supported by the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.B. (fbezani@ucla.edu).

Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock

Chanseok Shin[†], Ying Feng[†] & James L. Manley

Department of Biological Sciences, Columbia University, New York, New York 10027, USA

[†] These authors contributed equally to this work

The cellular response to stresses such as heat shock involves changes in gene expression¹. It is well known that the splicing of messenger RNA precursors is generally repressed on heat shock^{2,3}, but the factors responsible have not been identified^{4–8}. SRp38 is an SR protein splicing factor^{9,10} that functions as a general repressor of splicing. It is activated by dephosphorylation and required for splicing repression in M-phase cells¹¹. Here we show that SRp38 is also dephosphorylated on heat shock and that this dephosphorylation correlates with splicing inhibition. Notably, depletion of SRp38 from heat-shocked cell extracts derepresses splicing, and adding back dephosphorylated SRp38 specifically restores inhibition. We further show that dephosphorylated SRp38 interacts with a U1 small nuclear ribonucleoprotein particle (snRNP) protein, and that this interaction interferes with 5'-splice-site recognition by the U1 snRNP. Finally, SRp38-deficient DT40 cells show an altered cell-cycle profile consistent with a mitotic defect; they are also temperature sensitive and defective in recovery after heat shock. SRp38 thus plays a crucial role in cell survival under stress conditions by inhibiting the splicing machinery.

To examine whether the splicing repressor SRp38 functions in the inhibition of splicing that occurs after heat shock, we first examined whether the phosphorylation status of SRp38 changes in response to heat shock. SRp38 was efficiently dephosphorylated (dSRp38) when HeLa cells were incubated at 43 °C for durations of up to 1 h, whereas the phosphorylation status of classical SR proteins changed only modestly (Fig. 1a). The pattern of SRp38 dephosphorylation was distinct from that observed in mitosis¹¹ because, in addition to the fully dephosphorylated isoform, a partially dephosphorylated form accumulated in heat-shocked cells. Incubation of HeLa cells at higher temperatures (such as 46 °C) resulted in a much faster appearance of dSRp38 (data not shown).

We next tested whether SRp38 dephosphorylation could be induced by heat treatment of HeLa nuclear extract. This would allow us to rule out the possibility that the accumulation of dSRp38 requires new protein synthesis and/or the activity of classical heat-shock proteins, which are excluded from the nucleus before heat shock^{12,13}. Standard HeLa nuclear extract was incubated at 46 °C for 5–25 min and used to measure the phosphorylation status of SRp38 and splicing proficiency (at 30 °C). Splicing was rapidly and completely inhibited, and SRp38 was equally rapidly dephosphorylated (Fig. 1b, c).

We also 'heat-shocked' nuclear extract (46 °C for 5 min) and compared splicing and SRp38 phosphorylation over time with untreated extracts. A typical splicing profile was observed in the control extract, whereas almost complete inhibition was observed in the heat-shocked extract (Fig. 1d). Notably, SRp38 seemed to be fully phosphorylated in the control extract throughout the incubation, but was considerably dephosphorylated in the heat-shocked extract; in fact, dephosphorylation increased slightly during the time course (Fig. 1e). The phosphorylation status of another SR protein, ASF/SF2, did not change detectably in either extract. We also examined splicing and SRp38 phosphorylation status in

nuclear extract and lysate prepared from HeLa cells that were heat-shocked *in vivo* and then allowed to recover. SRp38 was completely converted to faster migrating forms immediately after heat shock, but then recovered such that at 60 min only the fully phosphorylated form was detected (Fig. 1f). Consistent with previous studies^{3–5}, nuclear extract from the heat-shocked cells was inactive, but was fully active after 60 min of recovery (Fig. 1g).

We next wanted to obtain direct evidence that SRp38 functions in heat-shock splicing repression. Using RNA affinity^{11,14}, we developed conditions that allowed efficient and selective depletion of SRp38 proteins from nuclear extract prepared from heat-shocked cells (see Methods). All SRp38 isoforms were specifically depleted without the depletion of other SR proteins (Fig. 2a). Notably, efficient splicing activity was restored to the SRp38-depleted extract (Fig. 2b), indicating that one of the SRp38 isoforms may be essential for splicing inhibition. The addition of 2–4 ng of dSRp38 to the depleted extract restored splicing inhibition, whereas larger amounts of phosphorylated SRp38, SC35 or ASF/SF2 had no effect (Fig. 2c). Addition of SC35, ASF/SF2 or total HeLa SR proteins to heat-shocked nuclear extract did not activate splicing (data not

shown), supporting the idea that standard SR proteins are not inactivated in heat-shocked cells^{6,7}, which is also the case in mitotic splicing repression¹¹. Together, our data indicate that SRp38 is rapidly dephosphorylated after heat shock and is responsible for the ensuing general inhibition of splicing.

To determine whether the inhibition of splicing that we detected after briefly heating HeLa nuclear extract also required SRp38, we depleted SRp38 from untreated HeLa nuclear extract by RNA affinity and measured the splicing competency of the extract before and after a 5-min incubation at 46 °C (Fig. 2d). No difference was observed between a mock and depleted extract in the absence of heat shock, consistent with the absence of significant amounts of dSRp38 in these extracts. Strikingly, however, splicing in the heated extracts was efficiently restored by SRp38 depletion. These results indicate that depletion of SRp38 from splicing extracts either after heat shock *in vivo* or before heat shock *in vitro* relieves splicing inhibition. Notably, however, splicing in the rescued extracts resulted in an accumulation of splicing intermediates. This was especially apparent in the extract heated *in vitro* (Fig. 2d), but also detectable in the extract heat-shocked *in vivo* (Fig. 2b), and is indicative of a

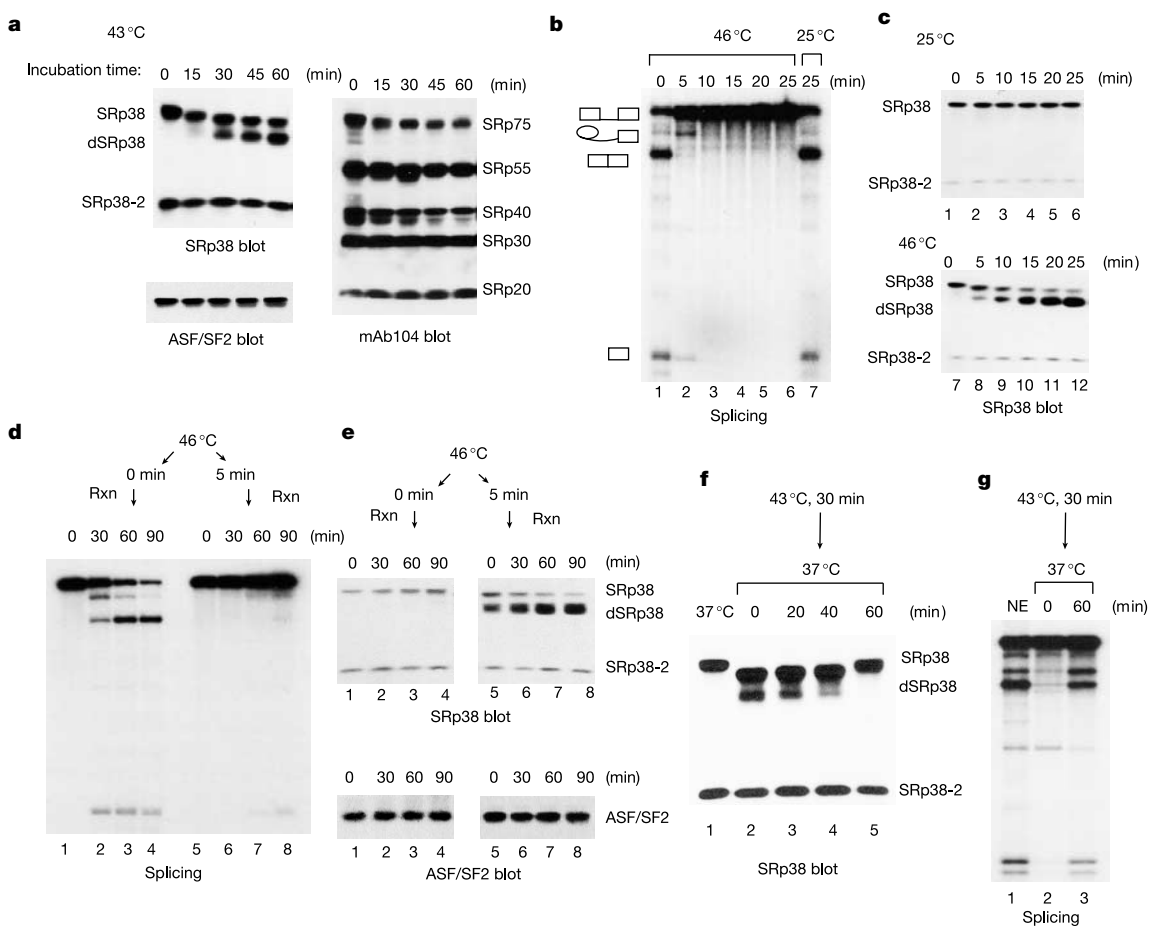


Figure 1 SRp38 is dephosphorylated and splicing is inhibited in response to heat shock. **a**, HeLa cells were incubated for the indicated times at 43 °C, and then collected and analysed by western blotting with the indicated antibodies. SRp38-2 is an isoform that arises from alternative splicing. **b**, Nuclear extract was incubated at 46 °C or 25 °C for the indicated times and assayed for splicing activity with β -globin pre-mRNA. **c**, Control (lanes 1–6) and heat-treated (lanes 7–12) nuclear extracts were analysed by western blotting with anti-SRp38 antibodies. **d**, HeLa nuclear extract was incubated at 46 °C for 0 or 5 min and splicing assays were carried out for the indicated times. **e**, Aliquots of the

extracts in **d** were analysed by western blotting with anti-SRp38 and anti-ASF/SF2 antibodies. **f**, HeLa cells were incubated at 43 °C for 30 min (lanes 2–5), and then at 37 °C for the indicated times. Non-treated cells were used for comparison (lane 1). Whole-cell lysates were analysed by western blotting with anti-SRp38 antibodies. **g**, Nuclear extracts were prepared from heat-treated cells either immediately or after 60 min of recovery (lanes 1, 2), and splicing activity was measured as in **b**. Standard nuclear extract was used as a control (lane 3).

defect in the splicing machinery that affects the second step of splicing. These results show directly that components of the splicing machinery may be heat sensitive and that splicing inhibition by dSRp38 helps to prevent possible mistakes.

We next investigated the mechanism by which dSRp38 (and not SRp38) inhibits splicing. We previously showed that changes in the phosphorylation status of RS domains of SR proteins can affect protein–protein interactions^{15,16}. Purified glutathione *S*-transferase (GST)-tagged SRp38 and dSRp38 were used in GST ‘pull-down’ assays (see Methods). GST–SRp38 but not GST–dSRp38 was found to interact very strongly with the splicing activator Tra2 α (Fig. 3a, left). We originally detected¹¹ SRp38 in a yeast two-hybrid screen with Tra2 α , and suspect that this interaction reflects an unknown function of SRp38 that is distinct from the repressor function of dSRp38. In sharp contrast, the U1 snRNP 70K protein (U1-70K) bound strongly to GST–dSRp38 but weakly to GST–SRp38 (Fig. 3a, right). This interaction seems to be very strong as it was completely resistant to 1 M NaCl (Fig. 3b). The association also occurs between endogenous proteins and reflects conditions in heat-shocked cells, as U1-70K was co-precipitated with SRp38 from nuclear extract prepared from heat-shocked cells but was co-precipitated much less efficiently from normal nuclear extract (Fig. 3c). The interaction with U1-70K is indicative of an association with U1 snRNP, as GST pull-down assays with nuclear extract assayed by northern blotting with small nuclear RNA (snRNA) probes showed preferential binding of GST–dSRp38 relative to GST–SRp38 and GST–ASF/SF2 (Fig. 3d). This assay detected interactions with other splicing snRNPs in addition to U1 snRNP, although it is unclear whether these interactions were direct or indirect.

The positive-acting SR protein ASF/SF2 stabilizes 5'-splice-site recognition by purified U1 snRNP through an RS-domain-mediated interaction between ASF/SF2 and U1-70K (ref. 17). To determine whether dSRp38 interferes with 5'-splice-site recognition, we carried out gel-shift assays with purified U1 snRNP, ASF/SF2 and the β -globin precursor mRNA (pre-mRNA) used in splicing (Fig. 3e). The presence of ASF/SF2 strongly stimulated formation of a stable (heparin-resistant) complex of U1 snRNP and RNA (Fig. 3e, lanes 1–3). The addition of either dSRp38 or SRp38 together with the other components had no significant effect (lanes 4–6). When U1 snRNP and either of the SRp38 proteins were briefly pre-incubated (4 °C, 5 min), however, dSRp38 but not SRp38 resulted in a strong, concentration-dependent decrease in complex formation (lanes 7–9). These results suggest that dSRp38 represses splicing by interacting tightly with the U1-70K protein, thereby interfering with the U1 snRNP 5'-splice-site interaction.

The above data and previous work¹¹ show that dSRp38 is a repressor of the splicing machinery in both mitotic and heat-shocked cell extracts. If these properties are physiologically important, then cells with reduced amounts of SRp38 should show defects related to the cell cycle and/or the stress response. We therefore generated chicken DT40 cells, which we previously used to study properties of other SR protein splicing factors^{18,19}, that lacked SRp38. Chickens contain one copy of *SRp38*, and the human and chicken proteins are 97% identical (data not shown). We constructed cell lines lacking either one (*SRp38*^{+/-}) or both (*SRp38*^{-/-}) alleles of *SRp38* (see Methods). Although SRp38 is not essential for cell viability, the SRp38-deficient cells showed an altered cell-cycle profile (Fig. 4a). Notably, G2/M phase was prolonged, most significantly in *SRp38*^{-/-} cells, but also in *SRp38*^{+/-} cells. These findings indicate that the mitotic splicing repressor function of SRp38 (ref. 11) may be important for proper mitotic progression.

To determine whether reduced amounts of SRp38 affect the response of DT40 cells to heat, we first examined whether increased temperatures induce SRp38 dephosphorylation in DT40 cells. Incubation at 44 °C for 1 or 2 h did not affect SRp38 phosphoryl-

ation (Fig. 4b). By contrast, when DT40 cells were exposed to 45.5 °C for a similar duration, SRp38 was efficiently dephosphorylated, showing a pattern of partial and more complete dephosphorylation that was essentially identical to that observed in HeLa cells (Fig. 4b). The pattern was qualitatively identical in wild-type and *SRp38*^{+/-} cells, but substantially less protein accumulated in the heterozygous cells.

To examine the possible significance of this induction of dSRp38, we measured the ability of these cell lines to recover after heat shock at 45.5 °C (Fig. 4c). Whereas the wild-type cells resumed rapid growth after about a 2-d lag, both the *SRp38*^{+/-} and *SRp38*^{-/-} cells showed considerably delayed recovery, resuming growth only after at least a 3-d lag. The *SRp38*^{-/-} cells recovered only marginally slower than the heterozygous cells, indicating that wild-type amounts of SRp38 are required for protection from the deleterious effects of heat shock (see below).

Although SRp38 dephosphorylation was not induced by 1–2 h of incubation at 44 °C, we wanted to determine whether SRp38 might be important for cell growth at increased temperatures. Previous studies^{20,21} have indicated that DT40 cells can grow for some time at about 43 °C. We measured the growth rate of wild-type, *SRp38*^{+/-} and *SRp38*^{-/-} cells at 44 °C (Fig. 4d). Whereas wild-type cells grew well for 2–3 d (but then ceased growing and eventually died),

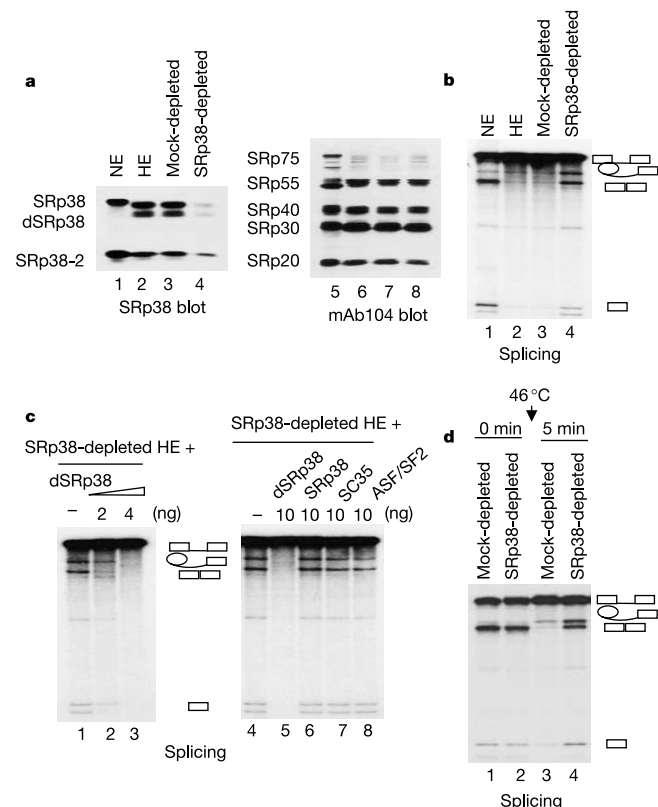


Figure 2 dSRp38 is necessary for heat-shock splicing repression. **a**, Biotinylated bio-C3 RNA was used to deplete SRp38 from nuclear extract prepared from heat-shocked cells. Standard nuclear extract (lanes 1, 5) and nuclear extract from heat-shocked cells depleted with bio-C3 RNA (lanes 4, 8), mock-depleted (lanes 3, 7) or untreated (lanes 2, 6) were analysed by Western blotting with the indicated antibodies. **b**, Extracts from **a** were tested for splicing activity using β -globin pre-mRNA. **c**, SRp38-depleted heat-shocked nuclear extract was supplemented with the indicated recombinant proteins and analysed for splicing activity. **d**, Normal HeLa extract was mock-depleted or depleted with bio-C3 RNA as in **a** and assayed for splicing without (lanes 1, 2) or with (lanes 3, 4) prior heating to 46 °C for 5 min.

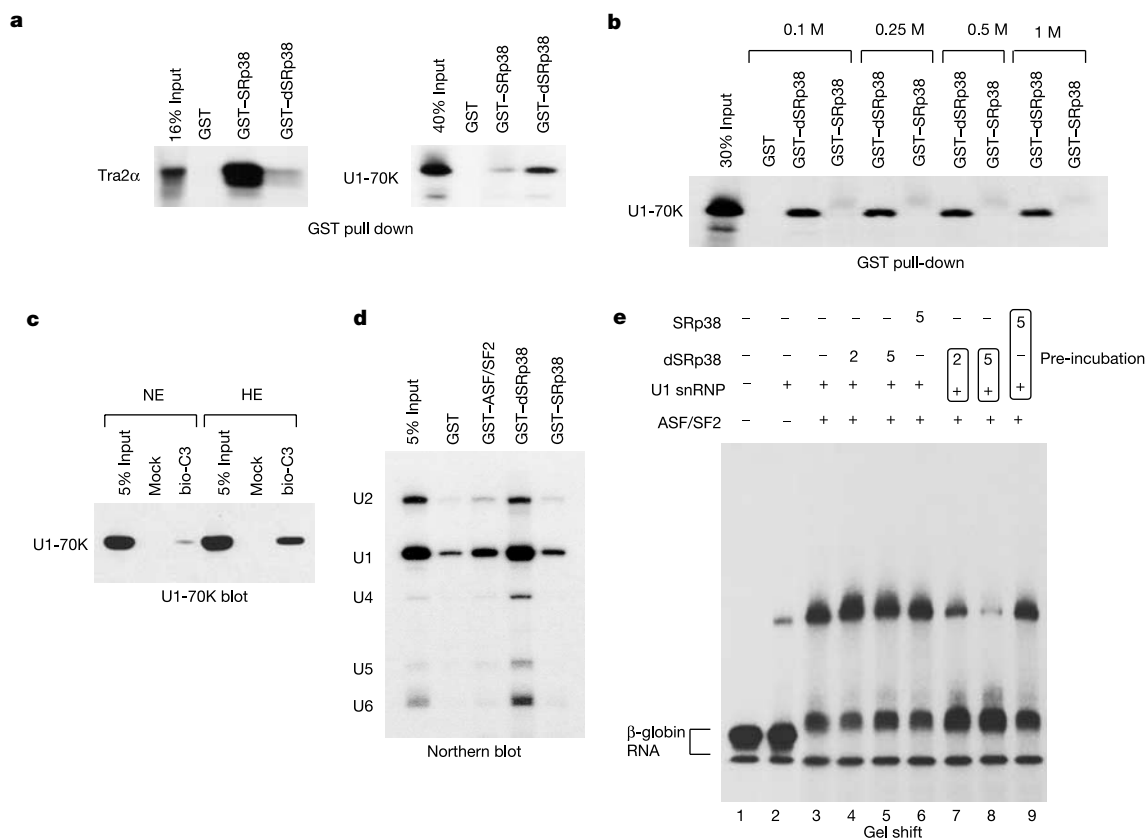


Figure 3 dSRp38 interacts with U1 snRNP and inhibits 5'-splice-site recognition. **a**, GST, GST-dSRp38 and GST-SRp38 proteins were used in binding reactions with [³⁵S]methionine-labelled Tra2α or U1-70K proteins. Reaction mixtures were treated with RNase and proteins were analysed by SDS-PAGE. **b**, GST fusion proteins were used in binding reactions with U1-70K as in **a**, except that bound proteins were washed with buffer containing the indicated concentration of NaCl. **c**, HeLa nuclear extract or nuclear extract prepared from heat-shocked cells was incubated with avidin beads (mock) or bio-C3 RNA beads in buffer containing 0.6 M KCl. Bound proteins were analysed by

western blotting with anti-U1-70K antibodies. **d**, GST fusion proteins were incubated with HeLa nuclear extract for 1 h at 30 °C. Eluted RNAs were purified and analysed by northern blotting with probes for U1, U2, U4, U5 and U6 snRNA. **e**, Purified U1 snRNP, ASF/SF2 and the indicated amount of SRp38 or dSRp38 (in ng) were incubated with ³²P-labelled β-globin pre-mRNA. The boxed combinations were pre-incubated for 5 min on ice before mixing with ASF/SF2 and RNA. After incubation, protein-RNA complexes were analysed by non-denaturing PAGE.

SRp38^{+/-} cells did not grow at all (although they remained viable) and SRp38^{-/-} cells began to die after about 1 d (such that less than 10% were viable after 60 h). Dephosphorylation of SRp38 was probably involved in maintaining cell survival because dSRp38 was detectable after about 8 h at 44 °C (Fig. 4e).

These results reflect loss of SRp38 rather than an unrelated property of the cells because, first, another independent SRp38^{-/-} cell line behaved identically to the first (data not shown); and second, stable expression of exogenous SRp38 (flu-SRp38) from a constitutive promoter considerably rescued the heat-shock recovery (Fig. 4f) and growth (data not shown) phenotypes. Rescue was incomplete, which probably reflects the fact that flu-SRp38, although efficiently dephosphorylated, was highly overexpressed in the rescued cell line (Fig. 4g). In support of this, a wild-type cell line expressing comparable amounts of flu-SRp38 (Fig. 4g) showed a phenotype nearly identical to the rescued cell line (see Fig. 4f).

Our data have shown that the SR protein SRp38 is crucial for cell survival after heat shock, reflecting its ability to inhibit the splicing machinery following dephosphorylation. The importance of this inhibition is highlighted by its evolutionary conservation from yeast, where splicing is rare, to human, where most genes are interrupted by intervening sequences. We have also shown that the splicing machinery is functional but suboptimal after heat shock

in the absence of SRp38. Thus, SRp38 probably represses splicing after heat shock to prevent, at least in part, the possible accumulation of inaccurately spliced mRNAs until the heat-damaged splicing apparatus is restored to normal, perhaps with the assistance of heat-shock proteins^{2,22}. SRp38 is highly conserved in vertebrates and may function as a heat-shock sensor and negative splicing regulator throughout vertebrates. An SR protein with related properties, RSF1 (ref. 23), may have a similar role in flies. As SR proteins do not seem to exist in yeast, another shut-off mechanism must apply in this organism.

Several studies³⁻⁵ have suggested that the integrity of snRNPs, especially the U4-U5-U6 complex, is compromised after heat shock. Our data indicate, however, that the splicing machinery remains largely functional provided that SRp38 is removed. It is possible that the reported defects in snRNPs reflect the activity of dSRp38, which, as we have shown, functions through a direct interaction with U1 and possibly other snRNPs. It may be that a defect in the U4-U5-U6 snRNP underlies the second-step splicing defect that we observed in the absence of SRp38. It may be important that the basic splicing machinery remains intact to allow the processing of intron-containing transcripts that must be expressed during heat shock; for example, some heat-shock protein mRNAs are derived from intron-containing precursors^{2,3}. In any event, our data, together with previous work establishing the role of

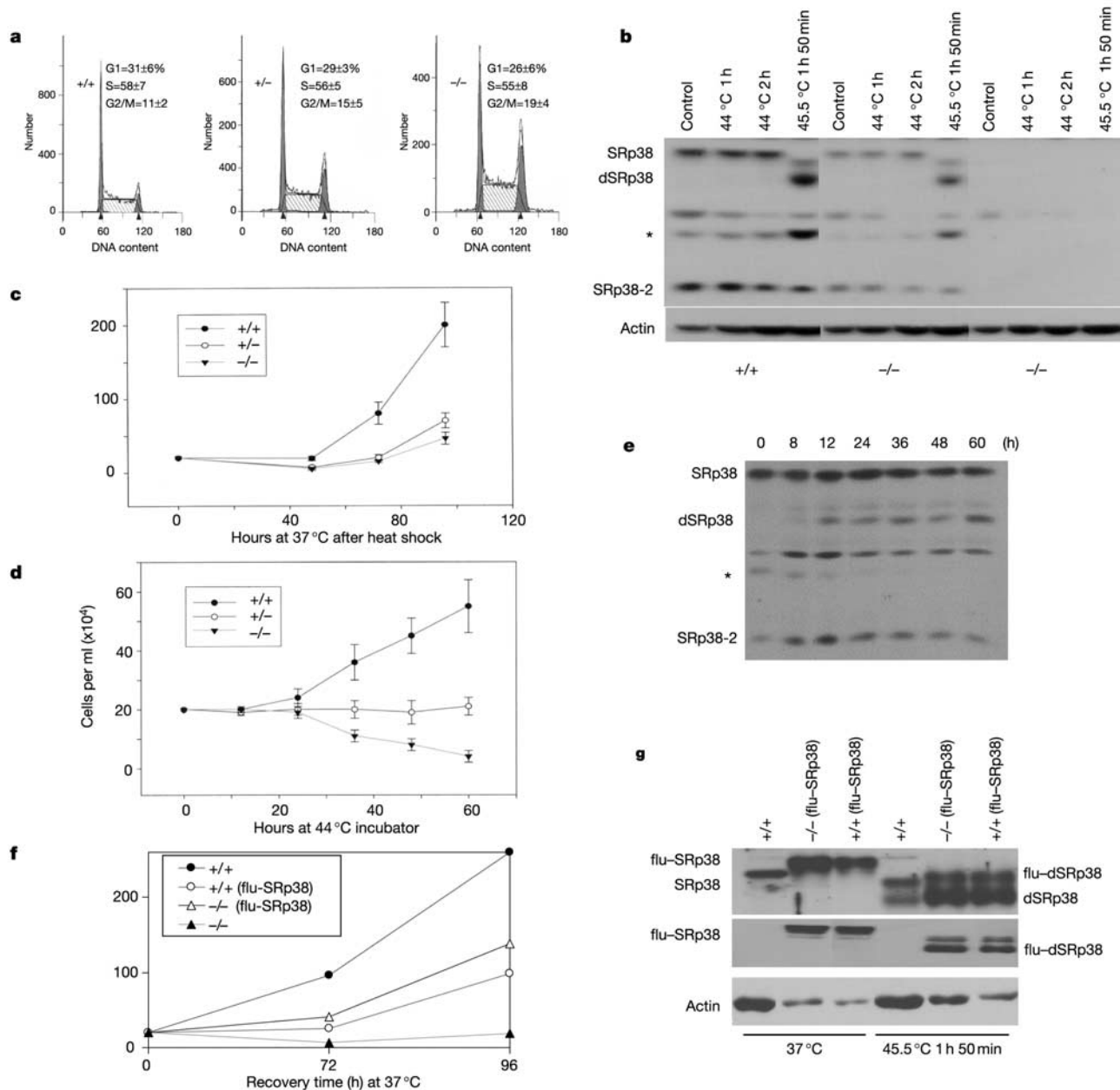


Figure 4 Characteristics of SRp38-deficient DT40 cells. **a**, Wild type (+/+), SRp38-heterozygous (+/-) and SRp38-deficient (-/-) cells were analysed by FACS. The percentage of cells in G1, S and G2/M is indicated. **b**, Cells were subjected to heat shock for the indicated times and at the indicated temperature. Western blot analysis of whole-cell lysates was done with anti-SRp38 or anti-actin antibodies. Asterisk represents a probable SRp38 breakdown product. **c**, Cells were heat-shocked at 45.5 °C for 110 min and allowed to recover at 37 °C. Numbers of surviving cells were determined by trypan

blue exclusion. Error bars represent the s.d. **d**, Cells were incubated at 44 °C and viability was determined as in **c**. Error bars represent the s.d. **e**, Cells were incubated at 44 °C for the indicated times. Western blot analysis of whole-cell lysates was done with anti-SRp38 antibodies. **f**, Cells expressing or not expressing exogenous flu-SRp38 were heat-shocked at 45.5 °C for 110 min and allowed to recover at 37 °C. Numbers of surviving cells were determined as in **c**. **g**, Western blot analysis of whole-cell lysates was done with anti-SRp38 (top), anti-flu (middle) or anti-actin (bottom) antibodies.

SRp38 in splicing repression during M phase, implicate SRp38 as a key global regulator of gene expression. □

Methods

Cell culture and heat treatment of cells or nuclear extract

HeLa cell culture and heat shock were done as described²⁴. In brief, monolayer cultures of HeLa cells at 80% confluency or HeLa nuclear extracts were immersed in a water bath for the indicated time and temperature. For recovery after heat shock, HeLa cells were incubated at 43 °C in a water bath for 30 min and moved to a 37 °C incubator for the indicated time immediately after changing the medium. We collected cells and washed them with cold PBS before western blotting or preparing nuclear extract for splicing assays. DT40 cells were plated at a density of 2×10^5 per ml, incubated at 44 °C for up to

60 h, or heat-shocked in a water bath for the indicated time and temperature, collected, washed once with PBS, resuspended in fresh medium and allowed to recover at 37 °C. Surviving cell numbers were determined by trypan blue exclusion.

Protein interaction assays

GST-SRp38 and GST-ASF/SF2 were purified from recombinant baculovirus-infected Sf9 cells by using glutathione-Sepharose 4B. Preparation of GST-dSRp38 with CIP was as described¹¹. Purified GST proteins (2 µg each per 10 µl of beads) were used for pull-down assays. [³⁵S]methionine-labelled U1-70K and Tra2α proteins were produced by *in vitro* translation using TNT rabbit reticulocyte lysate (Promega). Binding assays, including treatment with RNase A, were done essentially as described¹⁵. To analyse interactions with U snRNPs, glutathione beads containing GST fusion proteins were rotated with 75 µg of whole-cell extract from HeLa cells (extracted in 2 M NaCl) in buffer D as described¹⁵. After being washed in buffer D containing 250 mM NaCl, bound fractions were treated with

proteinase K (50 $\mu\text{g ml}^{-1}$) at 37 °C for 30 min and subjected to phenol–chloroform extraction, and RNA was recovered by ethanol precipitation. RNAs were separated by 10% PAGE and analysed by northern blotting with [α - ^{32}P]GTP-labelled antisense U1, U2, U4, U5 and U6 snRNA probes²⁵.

Gel-shift assays

Gel-shift assays were done as described¹⁷ except that β -globin RNA was used in the splicing assays. Recombinant His-tagged ASF/SF2, SRp38 and dSRp38 were prepared from recombinant baculovirus-infected Sf9 cells. ASF/SF2 (240 ng), pre-mRNA (16 ng), U1 snRNP (600 ng), dSRp38 (2–5 ng) and SRp38 (5 ng) was incubated under splicing conditions, except that ATP and creatine phosphate were omitted. Where indicated, dSRp38 or SRp38 was pre-incubated with U1 snRNP at 4 °C for 5 min, and reaction mixtures were incubated at 30 °C for 5 min. Heparin was then added to a concentration of 0.5 mg ml⁻¹ and reaction mixtures were incubated at 30 °C for an additional 5 min. We loaded samples on 4% acrylamide:bisacrylamide (80:1) gels containing 50 mM Tris and 50 mM glycine, and electrophoresis was done at 14 V cm⁻¹ for 3 h.

Details of western blotting, SRp38 depletion, splicing and generation of SRp38-deficient cells are given in the Supplementary Information.

Received 11 September; accepted 11 December 2003; doi:10.1038/nature02288.

- Yost, H. J., Petersen, R. B. & Lindquist, S. RNA metabolism: strategies for regulation in the heat shock response. *Trends Genet.* **6**, 223–227 (1990).
- Yost, H. J. & Lindquist, S. RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. *Cell* **45**, 185–193 (1986).
- Bond, U. Heat shock but not other stress inducers leads to the disruption of a subset of snRNPs and inhibition of *in vitro* splicing in HeLa cells. *EMBO J.* **7**, 3509–3518 (1988).
- Shukla, R. R., Dominicki, Z., Zwierzyński, T. & Kole, R. Inactivation of splicing factors in HeLa cells subjected to heat shock. *J. Biol. Chem.* **265**, 20377–20383 (1990).
- Utans, U., Behrens, S. E., Luhrmann, R., Kole, R. & Kramer, A. A splicing factor that is inactivated during *in vivo* heat shock is functionally equivalent to the [U4/U6.U5] triple snRNP-specific proteins. *Genes Dev.* **6**, 631–641 (1992).
- Gattoni, R. *et al.* The human hnRNP-M proteins: structure and relation with early heat shock-induced splicing arrest and chromosome mapping. *Nucleic Acids Res.* **24**, 2535–2542 (1996).
- Mahe, D. *et al.* Cloning of human 2H9 heterogeneous nuclear ribonucleoproteins. Relation with splicing and early heat shock-induced splicing arrest. *J. Biol. Chem.* **272**, 1827–1836 (1997).
- Bond, U. & James, T. C. Dynamic changes in small nuclear ribonucleoproteins of heat-stressed and thermotolerant HeLa cells. *Int. J. Biochem. Cell Biol.* **32**, 643–656 (2000).
- Graveley, B. R. Sorting out the complexity of SR protein functions. *RNA* **6**, 1197–1211 (2000).
- Manley, J. M. & Tacke, R. SR proteins and splicing control. *Genes Dev.* **10**, 1569–1579 (1996).
- Shin, C. & Manley, J. L. The SR protein SRp38 represses splicing in M phase cells. *Cell* **111**, 407–417 (2002).
- Arrigo, A. P. & Ahmad-Zadeh, C. Immunofluorescence localization of a small heat shock protein (Hsp23) in salivary gland cells of *Drosophila melanogaster*. *Mol. Gen. Genet.* **184**, 73–79 (1981).
- Velazquez, J. M. & Lindquist, S. Hsp70: nuclear concentration during environmental stress and cytoplasmic storage during recovery. *Cell* **36**, 655–662 (1984).
- Caputi, M., Mayeda, A., Krainer, A. R. & Zahler, A. M. hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *EMBO J.* **18**, 4060–4067 (1999).
- Xiao, S. H. & Manley, J. L. Phosphorylation of the ASF/SF2 RS domain affects both protein–protein and protein–RNA interactions and is necessary for splicing. *Genes Dev.* **11**, 334–344 (1997).
- Xiao, S. H. & Manley, J. L. Phosphorylation–dephosphorylation differentially affects activities of splicing factor ASF/SF2. *EMBO J.* **17**, 6359–6367 (1998).
- Kohtz, J. D. Protein–protein interactions and 5'–splice-site recognition in mammalian mRNA precursors. *Nature* **368**, 119–124 (1994).
- Wang, J., Takagaki, Y. & Manley, J. L. Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes Dev.* **10**, 2588–2589 (1996).
- Wang, J., Xiao, S. H. & Manley, J. L. Genetic analysis of the SR protein ASF/SF2: interchangeability of RS domains and negative control of splicing. *Genes Dev.* **12**, 2222–2233 (1998).
- Fukagawa, T., Regnier, V. & Ikemura, T. Creation and characterization of temperature-sensitive CENP-C mutants in vertebrate cells. *Nucleic Acids Res.* **29**, 3796–3803 (2001).
- Nakai, A. & Ishikawa, T. Cell cycle transition under stress conditions controlled by vertebrate heat shock factors. *EMBO J.* **20**, 2885–2895 (2001).
- Vogel, J. L., Parsell, D. A. & Lindquist, S. Heat-shock proteins Hsp104 and Hsp70 reactivate mRNA splicing after heat inactivation. *Curr. Biol.* **5**, 306–317 (1995).
- Labourier, E. *et al.* Antagonism between RSF1 and SR proteins for both splice-site recognition *in vitro* and *Drosophila* development. *Genes Dev.* **13**, 740–753 (1999).
- Dubois, M. F. *et al.* Heat shock of HeLa cells inactivates a nuclear protein phosphatase specific for dephosphorylation of C-terminal domain of RNA polymerase II. *Nucleic Acids Res.* **27**, 1338–1344 (1999).
- Hall, K. B. & Konarska, M. M. The 5' splice site consensus RNA oligonucleotide induces assembly of U2/U4/U5/U6 small nuclear ribonucleoprotein complexes. *Proc. Natl Acad. Sci. USA* **89**, 10969–10973 (1992).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank Z. Chen, T. Kashima, S. Bush, Y. Shi and X. Li for plasmids and discussion; A. Norris and A. Doty for technical assistance; I. Boluk for help preparing the manuscript; R. Luhrmann and B. Kastner for purified U1 snRNP; and W. van Venrooij for antibodies against U1-70K. This work was supported by a grant from the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.L.M. (jlm2@columbia.edu).

Molecular engineering of a backwards-moving myosin motor

Georgios Tsiavaliaris^{1,2}, Setsuko Fujita-Becker² & Dietmar J. Manstein^{1,2}

¹Institut für Biophysikalische Chemie, Medizinische Hochschule Hannover, OE 4350, Carl-Neuberg-Strasse 1, D-30623 Hannover, Germany

²Abteilung Biophysik, Max-Planck-Institut für Medizinische Forschung, Jahnstrasse 29, D-69120 Heidelberg, Germany

All members of the diverse myosin superfamily have a highly conserved globular motor domain that contains the actin- and nucleotide-binding sites and produces force and movement^{1,2}. The light-chain-binding domain connects the motor domain to a variety of functionally specialized tail domains and amplifies small structural changes in the motor domain through rotation of a lever arm^{3,4}. Myosins move on polarized actin filaments either forwards to the barbed (+) or backwards to the pointed (–) end^{5,6}. Here, we describe the engineering of an artificial backwards-moving myosin from three pre-existing molecular building blocks. These blocks are: a forward-moving class I myosin motor domain, a directional inverter formed by a four-helix bundle segment of human guanylate-binding protein-1 and an artificial lever arm formed by two α -actinin repeats. Our results prove that reverse-direction movement of myosins can be achieved simply by rotating the direction of the lever arm 180°.

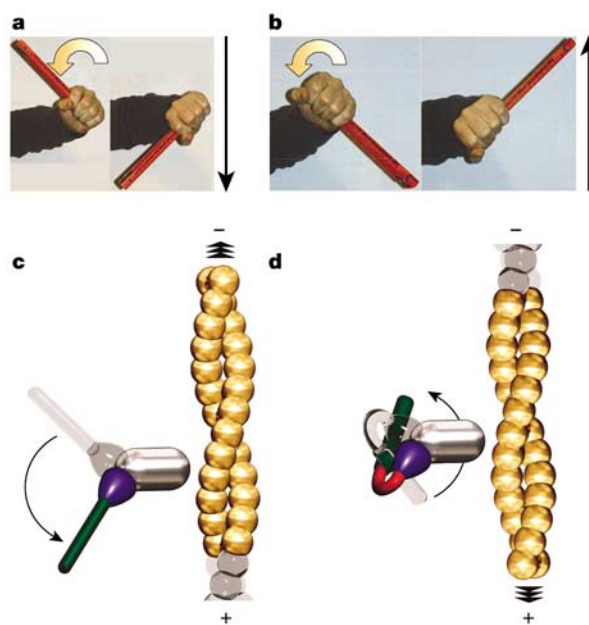


Figure 1 Mechanical models for myosin-based forwards and backwards movement. **a**, The translational movement of the tip of a lever (black arrow) depends on the angle of rotation and the direction in which the lever projects away from the axis of rotation. **b**, Simply by attaching the lever to the opposite site of the axis of rotation, the same rotation leads to a reversal of the translational movement of the lever. **c**, Power-stroke of a conventional barbed (+)-end-directed myosin. The converter region (blue) redirects the lever arm (green) such that it moves tangentially to the long axis of the actin filament (see also top views in Fig. 2). **d**, The insertion of an appropriately shaped domain (red) between the converter region and the lever arm reverses the direction of the power stroke and produces a pointed (–)-end-directed myosin. The myosin motor domain is shown in grey. The movement of the lever arm and force generation occur concomitantly with the release of ATP hydrolysis products. Additionally, myosin-based movement requires cyclic, ATP-dependent detachment and reattachment to actin filaments.

Most myosins move towards the barbed (+) end of actin filaments, but recent studies have established that at least one member of the family, myosin VI, moves towards the pointed (–) end⁵. The structural basis for reverse-direction movement has not been established. Two mechanisms for achieving reversal of myosin motility on the inherently polar actin filament have been suggested. On the basis of direct functional assays, electron microscopy and sequence analysis, Sweeney and co-workers proposed a model whereby reversal is achieved by rotating the lever arm in the opposite direction to conventional myosin lever arm movement⁵ (Fig. 1). Ikebe and co-workers, however, measured the motile properties of chimaeric constructs and proposed that the core of the motor domain is the sole determinant of directionality⁶.

To test whether rotating the lever arm 180° can itself produce reversal of directionality, we used molecular-engineering techniques. The constructs were based on the motor domain of the *Dictyostelium* class I myosin MyoE, which consists of 698 amino acids and will be referred to as E698. The structure of E698 is similar to that of conventional myosins⁷, but it lacks an amino-terminal SH3-like subdomain. The absence of this subdomain prevents steric clashes between the motor domain and the engineered lever arm. Additionally, E698 produces a larger power stroke by moving the attached lever arm through a larger angle than the wild-type form does^{7,8}. To rotate the lever arm in the opposite direction, we attached a four-helix bundle to the carboxy-terminal helix emerging from E698. The four-helix bundle comprises residues Leu 410 to Glu 508 of human guanylate-binding protein-1 (hGBP), the N-terminal tripeptide Arg-Glu-Met and a C-terminal Arg residue. The hGBP four-helix bundle, the structure of which has been determined to 1.8 Å resolution⁹, is an ideal building block for achieving rotation of the lever arm. It is a highly rigid structure whose orientation, after

fusion with the C-terminal helix of E698, can be accurately predicted. This fusion produces a forwards-stepping lever arm that is 4 nm in length and has the same orientation as the native LCBD (see Fig. 2).

A backwards-stepping lever arm 12 nm in length was created by the fusion of repeats 1 and 2 of α -actinin to the C-terminal helix of the hGBP four-helix bundle, which projects in the opposite direction to the helix emerging from the motor domain (see Supplementary movie 1). Previously, we have shown that the light-chain-binding domain (LCBD) of myosin II can be functionally replaced by building blocks consisting of two α -actinin repeats (2R) acting as an artificial lever arm^{10–12}. We have also solved the structure of the *Dictyostelium* myosin II motor domain fused to 2R (ref. 13). Therefore, the structures of all the building blocks used for the generation of the backwards-moving motor E698- Ω 2R are known, and they were connected by the direct fusion of α -helical regions of these building blocks. As a control, the motile activity of E698- Ω 2R was compared to that of E698-2R, a barbed (+)-end-directed motor consisting of the MyoE motor domain fused to two α -actinin repeats. Models of E698- Ω 2R and a myosin II motor domain fused to two α -actinin repeats (MD-2R) are shown in Fig. 2 bound to actin in the pre-power stroke state.

As the neck region of E698- Ω 2R contains both forwards- and backwards-acting lever arm elements, the generation of force and movement by E698- Ω 2R should depend on the motor's orientation during surface attachment. Unspecific attachment of E698- Ω 2R to the surface is predicted to produce no movement or only short runs of back- and forth-moving actin filaments in the *in vitro* motility assay. Continuous movement of actin filaments requires specific attachment through either the backwards- or forwards-acting lever arm element to the surface of the assay compartment. Our results

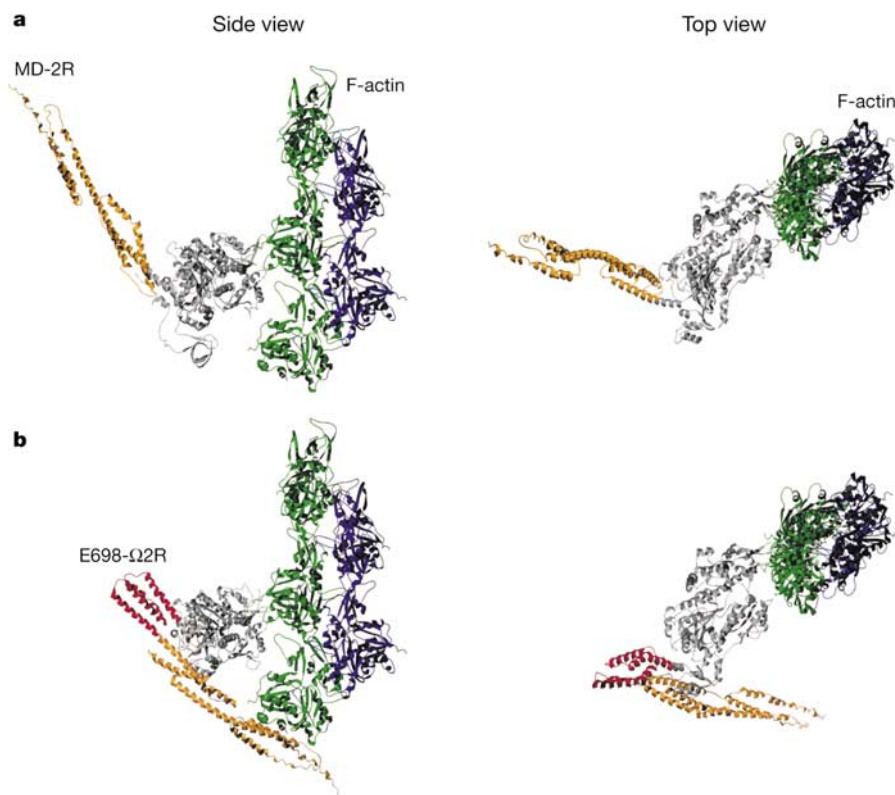


Figure 2 Molecular models of MD-2R and E698- Ω 2R attached to F-actin. **a**, The artificial barbed (+)-end-directed motor MD-2R consists of a myosin II motor domain (grey) and an artificial lever arm formed by two α -actinin repeats (orange). **b**, The artificial pointed (–)-end-directed motor E698- Ω 2R consists of the myosin I motor domain (grey), the

hGBP four-helix bundle (red) and the lever arm (orange). The motors are modelled in the 'pre-power-stroke' state attached to an actin protofilament consisting of five actin monomers (green and blue).

show that these predictions are correct. Short runs of back- and forth-moving actin filaments were observed on surfaces decorated with unspecifically oriented E698- Ω 2R (see Supplementary Fig. 1). The proportion of moving filaments decreased with increasing E698- Ω 2R density on the surface owing to the resulting shortening in run length. By contrast, unspecifically attached E698-2R supported continuous sliding of actin filaments with a velocity of $0.62 \pm 0.01 \mu\text{m s}^{-1}$ over a wide concentration range.

When the C-terminal His-tag, which is present in both E698- Ω 2R and E698-2R, was used to attach the constructs in a specific orientation to the anti-His-tag antibody-coated surface, both constructs moved actin filaments continuously and with velocities of approximately $0.7 \mu\text{m s}^{-1}$. Next, we determined the directionality of the motor constructs using polarity-marked actin filaments. Rhodamine/phalloidin was used to uniformly label and stabilize actin filaments, whereas Alexa Fluor 647 was used to label caps at the pointed (–) end. Rhodamine fluorescence is shown false-coloured in red and Alexa Fluor 647 fluorescence in green (Fig. 3a). Actin filaments moved with the pointed (–) end leading on surfaces that were coated either with E698-2R or with rabbit skeletal muscle heavy meromyosin (HMM). This result agrees with myosin-II-derived motors moving towards the barbed (+) end. By contrast, E698- Ω 2R attached to an anti-His-tag antibody-coated surface moved filaments with the pointed (–) end trailing (Fig. 3a), indicating that E698- Ω 2R is a pointed (–)-end-directed motor (see also Supplementary movies). The sliding velocities of labelled actin filaments obtained following their specific attachment are summarized in Fig. 3b. Positive velocity indicates movement towards the barbed (+) end, and negative velocity indicates movement towards the pointed (–) end of actin filaments. The wider

distribution of velocities observed with E698- Ω 2R results from a small but notable increase in the frequency of slow events. This bias is due to the presence of a small subpopulation of unspecifically attached E698- Ω 2R in addition to antibody-bound E698- Ω 2R.

The results of the *in vitro* motility assays confirm the reversal of the directionality of a barbed (+)-end-directed motor. The similar absolute velocities of E698- Ω 2R and E698-2R show that insertion of the hGBP four-helix bundle affects only the direction in which the lever arm projects away from the motor domain and not the protein's interactions with either nucleotides or actin. Our results agree with the findings of Sweeney and co-workers⁵ on the directionality of myosin VI. They support a model that proposes that myosins and kinesins, both sharing a common fold consisting of seven β -strands and six α -helices^{14,15}, are intrinsically plus-end-directed motors. Conformational changes in the core motor domain are either amplified or amplified and redirected by the neck region in both protein classes. In evolutionary terms, the insertion of a small domain is a more favourable way to change motor direction than modification of the core motor domain and the basic mechanism by which the motor interacts with nucleotides and its filamentous protein track.

This work builds on protein-engineering approaches developed by our laboratory to elucidate the mechanism underlying motor protein action^{16–22}. The control of critical motor parameters such as velocity, force and directionality should greatly facilitate future design and systematic characterization of novel motor proteins suitable for applications ranging from nano-technology to molecular medicine. The engineering of proteins with new and well-defined properties from known building blocks derived from biologically unrelated proteins has a wide range of applications. For example, we are combining fluorescent probe techniques with the engineered attachment of long amplifier elements to study the conformational dynamics of enzymes with high spatial and temporal resolution using near-field microscopy techniques. In addition, to change the dielectric properties of surfaces in a dynamic and tuneable manner, ordered arrays of molecular motors are being created on surfaces using molecular lithography and protein building blocks that support self-assembly into designed networks²³. □

Methods

Protein engineering and purification

The expression vector for the production of E698-2R contains sequences encoding a translational fusion of the first 698 residues of *Dictyostelium* myosin IE, residues 264–505 of *Dictyostelium* α -actinin, eYFP and a C-terminal (His)₈ tag. The expression vector for the production of E698- Ω 2R, in addition, contains the sequence encoding residues 410–508 of hGBP inserted between the myosin IE motor domain and the two α -actinin repeats. The mutation S336E was inserted into both constructs to mimic phosphorylation at the myosin TEDS site. For E698-2R, the sequence of the converter/2R fusion site reads LEMPRT/RASEQTK. For E698- Ω 2R, the junction between the motor domain and four-helix bundle reads RTREM/LLQ and between the four-helix bundle and 2R it reads HER/RASEQTK. Expression and purification of the artificial motors from *Dictyostelium* were performed as described previously¹⁶. Rabbit HMM was prepared according to ref. 24, and rabbit skeletal muscle actin was purified as described previously²⁵.

Model building and figure preparation

The three-dimensional models of the myosin motor domains attached to F-actin in the 'pre-power-stroke' state were obtained by fitting the atomic models of F-actin²⁶, the myosin II motor domain construct MD-2R (ref. 27), and E698- Ω 2R into the electron-density map of the *Dictyostelium* acto-S1 complex²⁸. Model building was done with the program Swiss-PdbViewer²⁹. Figures 1 and 2 were prepared using POV-Ray (Persistence of Vision Ray Tracer v3.02, 1997, <http://www.povray.org>).

Functional assays

Video-enhanced microscopy and analysis of actin filament sliding were performed as described previously³⁰. Double labelling of F-actin was performed according to ref. 6 with minor modifications. G-actin was labelled with a 20-fold molar excess of Alexa Fluor 647 maleimide in the presence of 100 mM KCl, 5 mM MgCl₂, 1 mM EGTA and 25 mM imidazole-HCl, pH 7.5 (buffer A) in the dark for 2 h at 4 °C. The reaction was stopped by the addition of a two-fold excess of β -mercaptoethanol. The resulting actin filament seeds (0.5–2 μm in length) were collected by centrifugation (100,000 g for 10 min). Labelled filaments were separated from residual dye by gel filtration (Sephadex G-25) in 5 mM Tris, 0.2 mM CaCl₂, 1 mM Na₃ at pH 7.5 (buffer B), and stabilized with phalloidin. For

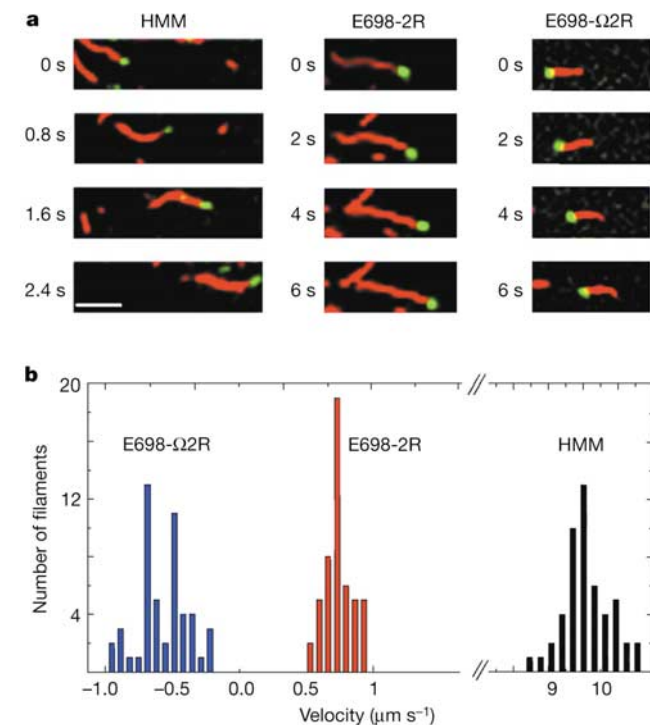


Figure 3 Direction of movement of myosin constructs. **a**, Movement of dual-labelled F-actin was visualized using a conventional *in vitro* motility assay²⁴. The green tip marks the pointed (–) end. HMM and E698-2R move filaments with their green tips leading; E698- Ω 2R moves them with their green tips trailing. Panels display the same spatial field at the relative times shown on the left. Bar, 10 μm . **b**, Histograms of the velocities of F-actin moving on surface-adsorbed HMM or on His-tag antibody-bound E698-2R and E698- Ω 2R. Movement towards the barbed (+) end is defined as positive movement.

Crystal structure and mechanism of a bacterial fluorinating enzyme

Changjiang Dong¹, Fanglu Huang², Hai Deng¹, Christoph Schaffrath¹, Jonathan B. Spencer², David O'Hagan¹ & James H. Naismith¹

¹Centre for Biomolecular Sciences, The University of St Andrews, Fife KY16 9ST, UK

²University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK

Fluorine is the thirteenth most abundant element in the earth's crust, but fluoride concentrations in surface water are low and fluorinated metabolites are extremely rare^{1,2}. The fluoride ion is a potent nucleophile in its desolvated state, but is tightly hydrated in water and effectively inert. Low availability and a lack of chemical reactivity have largely excluded fluoride from biochemistry: in particular, fluorine's high redox potential precludes the haloperoxidase-type mechanism^{3,4} used in the metabolic incorporation of chloride and bromide ions. But fluorinated chemicals are growing in industrial importance, with applications in pharmaceuticals, agrochemicals and materials products^{5–7}. Reactive fluorination reagents requiring specialist process technologies are needed in industry and, although biological catalysts for these processes are highly sought after, only one enzyme that can convert fluoride to organic fluorine has been described⁸. *Streptomyces cattleya* can form carbon–fluorine bonds⁹ and must therefore have evolved an enzyme able to overcome the chemical challenges of using aqueous fluoride. Here we report the sequence and three-dimensional structure of the first native fluorination enzyme, 5'-fluoro-5'-deoxyadenosine synthase, from this organism. Both substrate and products have been observed bound to the enzyme, enabling us to propose a nucleophilic substitution mechanism for this biological fluorination reaction.

When grown in the presence of F[–] ions, *S. cattleya* secretes fluoroacetate and 4-fluorothreonine, demonstrating its ability to biosynthesize organofluorine metabolites⁹. This organism contains an enzyme with a relative molecular mass (*M*_r) of 32,200 that has been shown to catalyse the formation of a C–F bond by combining S-adenosyl-L-methionine (SAM) and F[–] to generate 5'-fluoro-5'-deoxyadenosine (5'-FDA) and L-methionine (ref. 10 and Fig. 1). Purification¹¹ and now overexpression of 5'-fluoro-5'-deoxyadenosine synthase (5'-FDAS) have allowed a fuller characterization of activity: the enzyme has a catalytic rate constant (*k*_{cat}) of

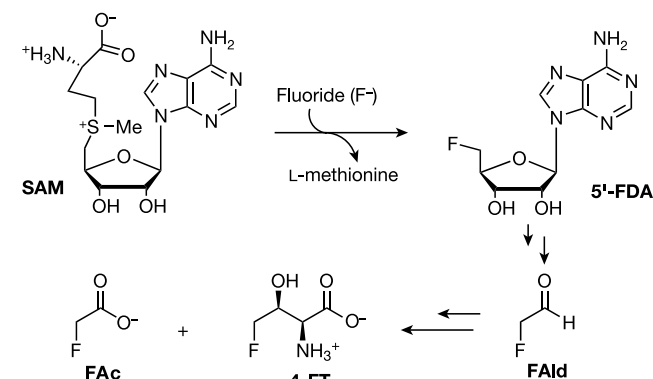


Figure 1 5'-FDAS from *S. cattleya* catalyses the formation of 5'-FDA from SAM and an F[–] ion. 5'-FDA is the first-formed organofluorine metabolite, which is ultimately converted to fluoroacetate (FAC) and 4-fluorothreonine (4-FT) through fluoroacetaldehyde (FAld) by *S. cattleya*²⁰. FAC is a toxin and 4-FT has antibiotic activity.

elongation, a 10-fold molar excess of unlabelled G-actin in buffer B containing 100 mM KCl was added to the Alexa-Fluor-647-labelled filaments and the reaction was allowed to proceed for 30–120 min on ice. Finally, the double-labelled filaments were diluted to a final concentration of 5 to 20 nM, and the elongated part of the filaments was labelled and stabilized by the addition of rhodamine–phalloidin. Double-labelled filaments were only used fresh and were discarded after 12 h.

Received 3 September; accepted 19 December 2003; doi:10.1038/nature02303.

1. Toyoshima, Y. Y. *et al.* Myosin subfragment-1 is sufficient to move actin filaments *in vitro*. *Nature* **328**, 536–539 (1987).
2. Manstein, D. J., Ruppel, K. M. & Spudich, J. A. Expression and characterization of a functional myosin head fragment in *Dictyostelium discoideum*. *Science* **246**, 656–658 (1989).
3. Rayment, I. *et al.* Structure of the actin-myosin complex and its implications for muscle contraction. *Science* **261**, 58–65 (1993).
4. Berg, J. S., Powell, B. C. & Cheney, R. E. A millennial myosin census. *Mol. Biol. Cell* **12**, 780–794 (2001).
5. Wells, A. L. *et al.* Myosin VI is an actin-based motor that moves backwards. *Nature* **401**, 505–508 (1999).
6. Homma, K., Yoshimura, M., Saito, J., Ikebe, R. & Ikebe, M. The core of the motor domain determines the direction of myosin movement. *Nature* **412**, 831–834 (2001).
7. Kollmar, M., Dürrewang, U., Kliche, W., Manstein, D. J. & Kull, F. J. Crystal structure of the motor domain of a class-I myosin. *EMBO J.* **21**, 2517–2525 (2002).
8. Veigel, C. *et al.* The motor protein myosin-I produces its working stroke in two steps. *Nature* **398**, 530–533 (1999).
9. Prakash, B., Renault, L., Praefcke, G. J., Herrmann, C. & Wittinghofer, A. Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. *EMBO J.* **19**, 4555–4564 (2000).
10. Uyeda, T. Q., Abramson, P. D. & Spudich, J. A. The neck region of the myosin motor domain acts as a lever arm to generate movement. *Proc. Natl Acad. Sci. USA* **93**, 4459–4464 (1996).
11. Anson, M., Geeves, M. A., Kurzawa, S. E. & Manstein, D. J. Myosin motors with artificial lever arms. *EMBO J.* **15**, 6069–6074 (1996).
12. Ruff, C., Furch, M., Brenner, B., Manstein, D. J. & Meyhöfer, E. Single-molecule tracking of myosins with genetically engineered amplifier domains. *Nature Struct. Biol.* **8**, 226–229 (2001).
13. Kliche, W., Fujita-Becker, S., Kollmar, M., Manstein, D. J. & Kull, F. J. Structure of a genetically engineered molecular motor. *EMBO J.* **20**, 40–46 (2001).
14. Kull, F. J. & Endow, S. A. Kinesin: switch I & II and the motor mechanism. *J. Cell Sci.* **115**, 15–23 (2002).
15. Vale, R. D. & Milligan, R. A. The way things move: looking under the hood of molecular motor proteins. *Science* **288**, 88–95 (2000).
16. Furch, M., Geeves, M. A. & Manstein, D. J. Modulation of actin affinity and actomyosin adenosine triphosphatase by charge changes in the myosin motor domain. *Biochemistry* **37**, 6317–6326 (1998).
17. Furch, M., Remmel, B., Geeves, M. A. & Manstein, D. J. Stabilization of the actomyosin complex by negative charges on myosin. *Biochemistry* **39**, 11602–11608 (2000).
18. Van Dijk, J. *et al.* Differences in the ionic interaction of actin with the motor domains of nonmuscle and muscle myosin II. *Eur. J. Biochem.* **260**, 672–683 (1999).
19. Knetsch, M. L. W., Uyeda, T. Q. & Manstein, D. J. Disturbed communication between actin- and nucleotide-binding sites in a myosin II with truncated 50/20-kDa junction. *J. Biol. Chem.* **274**, 20133–20138 (1999).
20. Tsiavaliaris, G. *et al.* Mutations in the relay loop region result in dominant-negative inhibition of myosin II function in *Dictyostelium*. *EMBO Rep.* **3**, 1099–1105 (2002).
21. Ponomarev, M. A., Furch, M., Levitsky, D. I. & Manstein, D. J. Charge changes in loop 2 affect the thermal unfolding of the myosin motor domain bound to F-actin. *Biochemistry* **39**, 4527–4532 (2000).
22. Reubold, T. F., Eschenburg, S., Becker, A., Kull, F. J. & Manstein, D. J. A structural model for actin-induced nucleotide release in myosin. *Nature Struct. Biol.* **10**, 826–830 (2003).
23. Ringler, P. & Schulz, G. E. Self-assembly of proteins into designed networks. *Science* **302**, 106–109 (2003).
24. Kron, S. J. & Spudich, J. A. Fluorescent actin filaments move on myosin fixed to a glass surface. *Proc. Natl Acad. Sci. USA* **83**, 6272–6276 (1986).
25. Lehrer, S. S. & Kerwar, G. Intrinsic fluorescence of actin. *Biochemistry* **11**, 1211–1217 (1972).
26. Lorenz, M., Poole, K. J., Popp, D., Rosenbaum, G. & Holmes, K. C. An atomic model of the unregulated thin filament obtained by X-ray fiber diffraction on oriented actin-tropomyosin gels. *J. Mol. Biol.* **246**, 108–119 (1995).
27. Dominguez, R., Frey, Y., Trybus, K. M. & Cohen, C. Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: visualization of the pre-power stroke state. *Cell* **94**, 559–571 (1998).
28. Schröder, R. R. *et al.* Three-dimensional atomic model of F-actin decorated with *Dictyostelium* myosin S1. *Nature* **364**, 171–174 (1993).
29. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).
30. Herm-Götz, A. *et al.* *Toxoplasma gondii* myosin A and its light chain: a fast, single-headed, plus-end-directed motor. *EMBO J.* **21**, 2149–2158 (2002).

Supplementary Information accompanies this paper on www.nature.com/nature.

Acknowledgements We thank S. Zimmermann for excellent technical assistance, R. Fedorov for providing Fig. 1, H. Faulstich for providing rhodaminephalloidin, C. Herrmann for the hGBP-1 cDNA, R. S. Goody, K. C. Holmes, M. A. Geeves, F. J. Kull, R. Maytum and D. P. Mulvihill for comments and discussions, and K. C. Holmes for continuous support. The work was supported by grants from the Deutsche Forschungsgemeinschaft (to D.J.M.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.M. (manstein@bpc.mh-hannover.de).

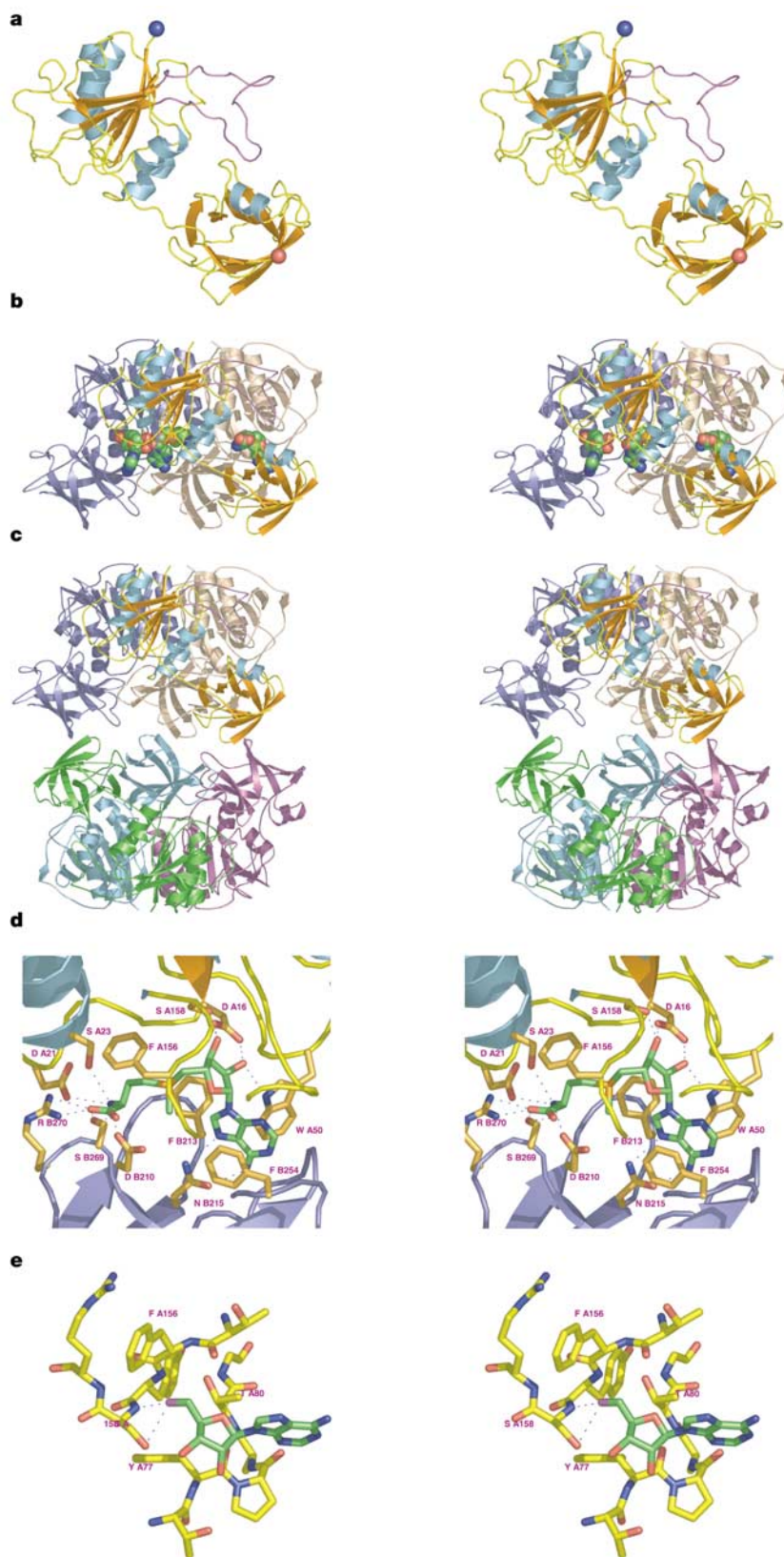


Figure 2 Structure of 5'-FDAS, drawn by PyMOL³¹. **a**, Stereoview of monomer unit (monomer A) in cartoon form, showing the N-terminal (top) and C-terminal (bottom) domains linked by a single polypeptide strand. The loop of 17 residues (98–114) unique to 5'-FDAS is shown in purple. **b**, The trimer formed by monomer A (coloured as in **a**), monomer B (blue) and monomer C (brown). Three molecules of SAM are shown in CPK format (carbon, green; oxygen, red; nitrogen, blue; sulphur, orange). The purple loop from the N-terminal domain sits in a cleft in the neighbouring N-terminal domain and in part forms the active site. SAM may act both as a substrate and in the structural integrity of the

protein. **c**, The hexameric structure of 5'-FDAS shown in smoothed cartoon form. 5'-FDAS is composed of a dimer of trimers. **d**, Stereoview of the substrate complex. The colour scheme is the same as in **b**. Monomers A and B are both involved in extensive recognition of the substrate, which is completely buried. There is very little perturbation in the location of the adenosine, ribose or methionine between the product and substrate complexes, and the interactions are conserved. **e**, Stereoview of the 5'-FDA complex. The colour scheme is the same as in **b** and fluorine is shown in mauve. The organofluorine makes a hydrogen bond with amide nitrogen and a polar contact with the O_γ of Ser 158.

0.07 min⁻¹, a Michaelis constant (K_m) for F⁻ of 2 mM, and a K_m for SAM of 74 μ M. Mutants of two glycosidase enzymes, which create a very potent electrophile, can catalyse carbon–halogen (F⁻, Cl⁻ and Br⁻) bond formation; however, the fluorinated compounds are reactive and have not been isolated⁸.

The gene encoding 5'-FDAS was cloned by degenerate polymerase chain reaction (PCR) on the basis of the partial protein sequence of the wild-type enzyme¹¹ using genomic DNA from *S. cattleya* as a template. Sequence information obtained from the degenerate PCRs was used to determine the complete coding sequence (*fIA*) for the enzyme. The *fIA* gene is 897 base pairs (bp) in size and encodes a protein containing 299 amino acids. The *fIA* gene amplified by PCR was inserted into the pET28a(+) plasmid and expressed in *Escherichia coli* BL21(DE3). The M_r of the purified recombinant product of *fIA* (Glu-Ser₂-His₆-Ser₂-Glu-Leu-Val-Pro-Arg-Glu-Ser-His-FlA) was 34,402, in agreement with electrospray ionization mass spectrometry (ESI-MS) data for the wild-type enzyme¹¹.

Crystallographic phases were determined by multiwavelength anomalous diffraction (MAD) from the selenomethionine (SeMet)-containing enzyme. These phases were used to determine the structure of the wild-type enzyme purified from *S. cattleya* to 1.9 Å. The refined structure contains residues 8–298 from the total of 299 residues: the first seven and the last residue are not visible. The monomer is organized as an amino-terminal (residues 8–180) and a smaller carboxy-terminal (residues 195–298) domain (Fig. 2a). A distinctive, long, extended loop of 15 residues connects the two domains, which are otherwise linked by few interactions. No structural homologues were found for either domain by the SSM¹² and the DALI¹³ queries, and thus there is no indication of an alternative or additional function.

The N-terminal domain has a central seven-stranded β -sheet, which combines parallel and antiparallel strands sandwiched between α -helices (Fig. 2a). The C-terminal domain is composed of a five- and a four-stranded antiparallel β -sheet. The Pfam database¹⁴ assigns the enzyme to family PF01887, and the structure provides the fold for this superfamily. Notably, no member of this superfamily has a known function. The asymmetric unit of the crystal contains a trimer (Fig. 2b): the three N-terminal domains are arranged around a three-fold axis and make intimate contacts with each other; the three C-terminal domains make contacts with their neighbouring N-terminal domain but have no contact with each other.

Sequence searching¹⁵ identifies several homologues from various bacteria and archaea. Sequence alignment of these homologues shows that an extended loop (98–114) in the N-terminal domain that is involved in the trimer contacts is missing in all homologues (Fig. 2a, b). In a model of the enzyme without this loop, the buried surface area is reduced by 25% and 8 of the 13 hydrogen bonds at

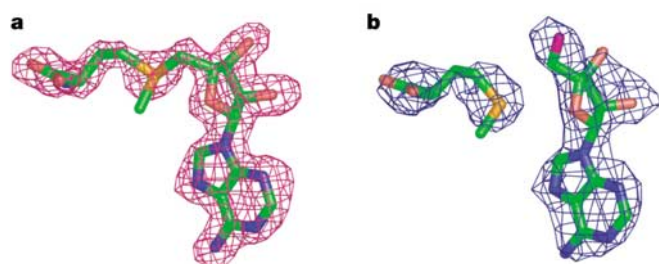


Figure 3 Fo–Fc electron density maps with phases calculated from models that do not include ligand. The colour scheme is the same as in Fig. 2. **a**, Map contoured at 3 σ for SAM, shown in magenta chicken wire. **b**, Map contoured at 2.6 σ for 5'-FDA and methionine, shown in blue chicken wire.

each of the three trimer interfaces are deleted. Thus, although the superfamily shares the same monomeric fold, 5'-FDAS may well have a unique quarternary structure. Crystallographic symmetry generates a hexamer (Fig. 2c), which is a dimer of the asymmetric unit trimer. Gel filtration indicates a hexamer¹¹.

The enzyme crystallized from *S. cattleya* contains a bound SAM molecule (Figs 2d and 3a), and SAM clearly does not dissociate readily from the protein during purification. SAM is bound at the interface between the C-terminal domain of one monomer and the N-terminal domain of the neighbouring monomer (Fig. 2b, d), and three molecules are bound by the trimer. The SAM molecule is completely buried by the protein on binding, which means that the protein must have an alternative open conformation during the turnover cycle. The recognition of SAM is highly specific (Figs 2d and 4). For ease of discussion, SAM can be decomposed into three components: the adenosine ring, the ribose ring and methionine. Each of these three components is recognized by a combination of hydrogen bonds and van der Waals contacts. All three components of SAM are recognized by both monomers, which indicates that the precise and possibly unique quarternary structure of the enzyme is crucial for substrate recognition and catalysis.

It seems reasonable that the extensive contacts between SAM and the protein would drive closure of the domains to form the enveloped binding site. Notably, SAM is bound in a high-energy conformation, and the C2–O2 and C3–O3 bonds of the ribose ring are found in an eclipsed conformation (torsion angle 1°). A search of the non-disordered well-refined entries in the small-molecule database¹⁶ with a furanose ring fragment query found only 1 ring from 264 possibilities that has a torsion angle for O2–C2–C3–O3 within 10° of the angle seen here. An analysis of well-refined high-resolution (<2.0 Å) protein structures¹⁷ shows that this torsion angle is always greater than 19° for the 15 SAM molecules in these structures. Molecular energy calculations indicate that this conformation relaxes on minimization with any of the common force fields (AMBER, MM2 and MM3)¹⁸.

Observations from isotopic labelling¹⁹ in whole-cell experiments indicate that the fluorination reaction occurs with a stereochemical 'inversion of configuration' at the C5' carbon of SAM to generate

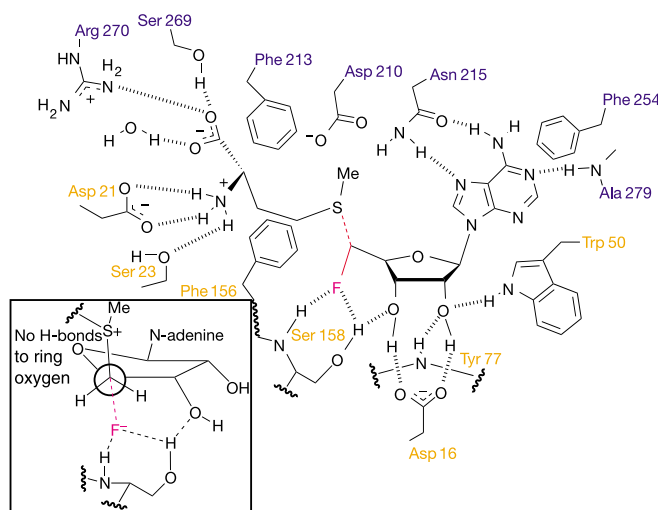


Figure 4 Representation of 5'-FDA and methionine bound to the enzyme, showing hydrogen-bonding to the fluoromethyl group from Ser 158, and the *anti* relationship between the C–F bond (red) and the disconnected C–S bond (dotted red) of SAM that is indicative of an S_N2 reaction course. Key residues are shown (monomer A, orange; monomer B, blue). Inset, the trajectory of the S_N2 and the conformation of ribose that minimizes negative stereoelectronic effects.

5'-FDA consistent with a nucleophilic substitution (S_N2) process. To define the molecular mechanism, we grew crystals of the enzyme with the two products of the reaction. This was done by incubating SeMet protein with 50 mM KF and 10 mM SAM for 4 h before crystallization. The resulting crystals diffracted to 2.7 Å and have essentially the same unit cell as the wild-type (SAM-bound) protein. The protein structure is identical to that observed for the SAM complex. It is immediately obvious from the electron density map that the SAM molecule is broken into two fragments (Fig. 3b). We built methionine and 5'-FDA into the density. The molecules are most clearly defined in the A subunit and reported distances relate to this subunit; however, the locations are essentially identical in all three subunits.

Methionine is bound in a manner very similar to that observed in the SAM complex; the only difference is in the torsion angles (maximum 16°) of the side chain, which results in a shift in the S and C ϵ atoms of 1.0 Å between the two structures. A similar observation is made when comparing 5'-FDA and SAM, where the positions of the adenosine and the ribose ring are unchanged. The planarity of the ribose ring refines to a different value in 5'-FDA, because the C2–O2 and C3–O3 bonds are no longer eclipsed; however, the resolution of this structure precludes detailed analysis. The C5' atom of the ribose ring in the 5'-FDA complex has shifted by 0.8 Å from its position in the SAM complex. The electron density clearly shows that the F atom is located in a pocket defined by a main chain stretch from Phe 156 to Ser 158 and the side chains of Phe 156, Tyr 77, Thr 80 and Ser 158 (Fig. 2e).

There is a clear hydrogen bond (N–F, 3.1 Å) between the amide NH of Ser 158 and the F atom of the fluoromethyl group, which also makes a second polar contact with the side chain of Ser 158. It is well known that organic, bound fluorine is a relatively poor hydrogen-bonding acceptor²⁰; however, the short contact here is very clear, suggesting an optimal NH–F contact of 2.0 Å (ref. 21). The pocket is unoccupied in the SAM complex and is hydrophobic in nature. There are no water molecules in the vicinity of the fluorine, which suggests that the F^- ion is fully dehydrated before C–F bond formation. The radius of this pocket is roughly 1.4–1.6 Å (assuming hydrogen atoms at their calculated positions with a van der Waals radius of 1.0 Å), which would preferentially select the F^- ion and possibly water over the other halogens. This would explain the observation that 5'-FDAS does not use other halogens.

Examination with the program HOLE²² does not find a convincing channel that is wide enough to allow the F^- anion to diffuse into the active site when SAM is bound, indicating that F^- may be bound first. Comparing the product and substrate complexes gives a distance between F (5'-FDA) and C5' (SAM) of 1.9 Å, aligns the C–F bond of the product (5'-FDA) *anti* (164°) to the C5'–S bond of the substrate (SAM) and reveals only very small structural changes between substrate and complex. We regard this as extremely strong structural evidence for an S_N2 mechanism (Fig. 4), with these structures representing the 'before and after'. This is consistent with previous stereochemical data¹⁹. Implicit in this is the requisite that the F atom of the fluoromethyl group in 5'-FDA is bound at the F^- -binding site and secured by the same hydrogen-bonding contacts used to anchor F^- .

It is our interpretation that the high-energy conformation of the ribose ring is optimal for the nucleophilic substitution reaction mediated by the enzyme. The pucker in the ring will act as a force on C5', weakening the C5'–S bond. The conformation forces a F^- trajectory orthogonal to the C4–O4 bond, minimizing electrostatic repulsion during the S_N2 reaction. The inductive influence of the ring oxygen will strengthen the C5'–S bond (Fig. 4). O4 is prevented, however, by the ribose conformation from involvement in an anomeric interaction with the adenine ring and has no hydrogen bonds. Both these factors would substantially increase the inductive influence of O4 hindering the reaction. We note that O3 of SAM is hydrogen-bonded to Ser 158 OH, which contacts F^- . Therefore, the position of O3 may be important for binding F^- .

SAM is extensively recognized, and its co-purification with the protein suggests that it is tightly bound. We predict that the SAM binding must drive the dehydration of F^- as it pushed into its binding pocket. Full desolvation of F^- would require roughly 400 kJ mol^{-1} : a very high activation energy for the enzyme. Extrapolating from the fluoromethyl of the product structure, it seems that F^- makes polar contacts with the enzyme through a strong hydrogen bond to the backbone amide and a bifurcated hydrogen bond to the serine OH. It is clear that the enzyme uses these contacts to avoid having to pay the full free energy of desolvation. Thus 5'-FDAS specifically dehydrates the F^- ion, making it a very potent nucleophile. The F^- ion is collinear with the C–S bond of SAM, which the enzyme forces into a conformation designed to promote nucleophilic cleavage of the C–S bond. Although no

Table 1 Experimental data

Data collection Wavelength (Å)	MAD (methionine and 5'-FDA)			SAMO.932
	Peak 0.9786	Inflection 0.9783	Remote 0.8984	
Resolution (highest shell, Å)	53–2.7 (2.74–2.67)	53.5–3.1 (3.18–3.10)	45.9–3.1 (3.18–3.10)	65–1.9 (1.98–1.90)
Space group			C222 ₁	
Cell constants (Å) $\alpha = \beta = \gamma = 90^\circ$		$a = 76.2, b = 129.8, c = 183.9$		$a = 75.9, b = 130.3, c = 183.4$
Unique reflections	26,303	17,635	16,925	67,204
Average redundancy	6.3 (3.5)	3.0 (2.8)	3.4 (3.3)	10 (7.7)
I/σ	8.5 (4.5)	11.9 (10.3)	14.1 (11.3)	4.8 (1.7)
Completeness (%)	97 (81)	84 (72)	82 (69)	94 (93)
Anom complete (%) [*]	97 (81)	84 (72)	82 (69)	–
$R_{\text{merge}}^\dagger$	6.4 (13)	4.2 (6.8)	3.7 (6.3)	11.0 (44.0)
Refinement				
R	17.6	–	–	16.7
R_{free}	23.9	–	–	21.7
r.m.s. deviation bonds (Å)/angles ($^\circ$)	0.014/1.34	–	–	0.017/1.65
B-factor deviation bonds/angles (Å ²):	0.51 / 0.87	–	–	1.15/1.8
main chain side chains	1.35/2.28	–	–	2.53/3.85
Residues in Ramachandran core (%)	84	–	–	88
Protein atoms	6,660	–	–	6,660
Water atoms	158	–	–	719
Ligand atoms	84	–	–	81
Average B-factor (Å ²)	23.1	–	–	18.4
PDB accession code	1RQR	–	–	1RQP

^{*}Anomalous completeness corresponds to the fraction of possible acentric reflections for which an anomalous difference has been measured.

[†] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i - \langle I \rangle| / \sum_{hkl} \langle I \rangle$, where I_i is an intensity for the i th measurement of a reflection with indices hkl , and $\langle I \rangle$ is the weighted mean of the reflection intensity.

structural data are available, in the most efficient glycosidase mutant capable of C–F bond formation, a hydrogen bond with serine has been proposed to direct a dehydrated F[−] ion⁸.

In summary, our data have revealed details of biological fluorination, whereby nature converts inorganic fluoride to organic fluorine, opening up an area of research focusing on the production of highly valuable organofluorine compounds by biotechnological means. □

Methods

Cloning and overexpression of FIA in *E. coli*

Three degenerate primers, pFN (5′-AAAAGGATCCGCGSCAACWSSACSCGSCGCC SATCATC-3′), PFI-1 (5′-AAAAGAATTCGCGYTRGAARCCYGCRCSSWRCSCGCCA-3′) and PFI-2 (5′-NCKNSWRTARTARAANGTNGGYTCNGGYTGYC-3′), were designed on the basis of the partial amino acid sequence of the wild-type enzyme¹¹ synthesized, and used for amplification of the corresponding gene fragments by PCR. All primers were synthesized by MWG-Biotech. The PCR products were subcloned into pUC18 and sequenced. We used two primers, pSCF1 (5′-TGACGTCCGGCAGATGCTG TACATGAGCCCCCTTGCACT-3′) and pSCF2 (5′-GAGGAGCACGGCTACCTGGAGGC GTACGAGGTTCACCTCG-3′), in further gene walking to identify the rest of the *fIA* gene sequence. *fIA* was amplified by PCR from genomic DNA of *S. cattleya* using the primers pSCF3 (5′-GCAGGAGGAATTCATATGGCTGCCAACAGC-3′) and pSCF4 (5′-CCCTACGCGCTCGAGGGTACGTCGTCGC-3′). Three bases were changed (indicated by italics) in the primers to create *Nde*I and *Xho*I restriction sites (underlined) for cloning. The ATG in the *Nde*I site is the start codon of the gene. The pSCF4 primer is located 50-bp downstream of the stop codon of the *fIA* gene. The PCR product was inserted into the pET28a(+) vector (Novagen) between the *Nde*I and *Xho*I sites to generate pET28-*fIA*. This plasmid directs the expression of FIA with a small His-tag-containing peptide (Met-Glu-Ser₂-His₆-Ser₂-Glu-Leu-Val-Pro-Arg-Glu-Ser-His) fused to the N-terminal of the enzyme.

E. coli BL21(DE3) transformed with pET28-*fIA* was grown in Luria broth containing 100 µg ml^{−1} kanamycin at 37 °C until an absorbance of 0.4–0.6 at 600 nm was reached. We induced overexpression of FIA by adding isopropylthiogalactoside (IPTG) to 0.2 mM and continued the incubation at 16 °C overnight. Cells were collected and lysed by sonication. After centrifugation, the cell lysate was applied onto a column packed with Ni²⁺-charged His-Bind resin (Novagen). Recombinant protein bound on the resin was eluted with 20 mM Tris-HCl (pH 7.9), 250 mM imidazole, 0.5 M NaCl and 10% glycerol. The protein was further purified by gel filtration fast protein liquid chromatography (FPLC) on a Superdex S-200 (HR16/60) column (PharmaciaBiotech) and concentrated using a Vivaspinn concentrator (Vivascience). We used Bradford reagent (BioRad) to measure protein concentrations. The *M_r* of the protein was determined by polyacrylamide gel electrophoresis, gel-filtration FPLC and ESI-MS analysis.

Assay of the recombinant fluorination enzyme

In a total reaction volume of 100 µl, 15 µg of FIA was incubated with 0.8 mM SAM, 10 mM NaF and 50 mM Tris-HCl (pH 7.9) at 37 °C for 3 h. The reaction product 5′-FDA was confirmed by liquid chromatography mass spectrometry (LC-MS) using synthetic 5′-FDA¹⁰ as a standard.

Crystallography

Full details of the purification¹¹ and crystallization of the enzyme from *S. cattleya* have been published elsewhere²³. The 1.9-Å data set was recorded on ID14-2 of the European Synchrotron Radiation Facility (ESRF). The protein crystals were thought to be wild type, but on structure elucidation it became clear that SAM was bound. SeMet-labelled protein was produced by overexpression in *E. coli* by the methionine inhibition protocol²⁴ and confirmed by mass spectroscopy. Crystals of 5′-FDA and methionine complex were obtained for SeMet protein under similar conditions to those used for the 'native' SAM crystals, except that the enzyme was first incubated with 50 mM KF and 10 mM SAM for 4 h.

Data were collected at three wavelengths on BM14 UK at the ESRF and reduced with MOSFLM/SCALA²⁵; full details are given in Table 1. SOLVE²⁶ located 18 Se positions and generated phases. Automatic interpretation of electron density was done by combining the phases from the SeMet crystals with the isomorphous 1.9 Å in Arp/wARP²⁷. This procedure led to an essentially complete chain trace of the monomer. Both structures were refined by using REFMAC²⁸ and statistics are given in Table 1. Dictionaries were derived from the PRODIG server²⁹; however, all torsion restraints were removed from the ribose ring to ensure a non-dictionary-biased conformation was obtained from refinement. Experimental data and structures have been deposited in the relevant data banks.

Received 2 October; accepted 9 December 2003; doi:10.1038/nature02280.

- O'Hagan, D. & Harper, D. B. Fluorine-containing natural products. *J. Chem.* **100**, 127–133 (1999).
- Xu, X.-H. *et al.* 5-Fluorouracil derivatives from the sponge *Phakellia fusca*. *J. Nat. Prod.* **66**, 285–288 (2003).
- vanPee, K. H. Biosynthesis of halogenated metabolites by bacteria. *Annu. Rev. Microbiol.* **50**, 375–399 (1996).
- Littlechild, J. Haloperoxidases and their role in biotransformation reactions. *Curr. Opin. Chem. Biol.* **3**, 28–34 (1999).
- Sandford, G. Organofluorine chemistry. *Phil. Trans. R. Soc. Lond. A* **358**, 455–471 (2000).
- Mann, J. Modern methods for the introduction of fluorine into organic molecules—an approach to compounds with altered chemical and biological activities. *Chem. Soc. Rev.* **16**, 381–436 (1987).
- Hutchinson, J. & Sandford, G. Elemental fluorine in organic chemistry. *Top. Curr. Chem.* **193**, 1–43 (1997).
- Zechel, D. L. *et al.* Enzymatic synthesis of carbon–fluorine bonds. *J. Am. Chem. Soc.* **123**, 4350–4351 (2001).
- Sanada, M. *et al.* Biosynthesis of fluorothreonine and fluoroacetic acid by the thienamycin producer, *Streptomyces cattleya*. *J. Antibiotics* **141**, 259–265 (1986).
- O'Hagan, D., Schaffrath, C., Cobb, S. L., Hamilton, J. T. G. & Murphy, C. D. Enzyme catalysed organofluorine synthesis. *Nature* **416**, 279 (2002).
- Schaffrath, C., Deng, H. & O'Hagan, D. Isolation and characterisation of 5′-fluoro-deoxyadenosine synthetase, a fluorination enzyme from *Streptomyces cattleya*. *FEBS Lett.* **547**, 111–114 (2003).
- Boutselakis, H. *et al.* E-MSD: the European Bioinformatics Institute Macromolecular Structure Database. *Nucleic Acids Res.* **31**, 458–462 (2003).
- Holm, L. & Sander, C. Protein folds and families: sequence and structure alignments. *Nucleic Acids Res.* **27**, 244–247 (1999).
- Bateman, A. *et al.* The Pfam Protein Families Database. *Nucleic Acids Res.* **30**, 276–280 (2002).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. B* **58**, 380–388 (2002).
- Berman, H. M. *et al.* The Protein Data Bank. *Acta Crystallogr. D* **58**, 899–907 (2002).
- Kollman, P. A. *et al.* Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897 (2000).
- O'Hagan, D. *et al.* An assay for the enantiomeric assay of [²H]₁-fluoroacetic acid: Insight into the stereochemical course of fluorination during fluorometabolite biosynthesis in *Streptomyces cattleya*. *J. Am. Chem. Soc.* **125**, 379–387 (2003).
- Dunitz, J. & Taylor, R. Organic fluorine hardly ever accepts hydrogen bonds. *Chem. Eur. J.* **3**, 89–98 (1997).
- Howard, J. A. K., Hoy, J. V., O'Hagan, D. & Smith, G. T. How good is fluorine as a hydrogen bond acceptor? *Tetrahedron* **38**, 12613–12622 (1996).
- Smart, O. S., Neduveil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. P. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph. Model* **14**, 354–360 (1996).
- Dong, C. *et al.* Crystallization and X-ray diffraction of the 5′-fluoro-5′-deoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya*. *Acta Crystallogr. D* **60**, 760–763 (2003).
- Double, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
- Bailey, S. The CCP4 Suite—programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr. D* **58**, 968–975 (2002).
- Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D* **55**, 247–255 (1999).
- van Aalten, D. M. F. *et al.* PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput. Aided Mol. Des.* **10**, 255–262 (1996).
- Schaffrath, C., Cobb, S. L. & O'Hagan, D. Cell-free biosynthesis of fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. *Angew. Chem. Int. Edn Engl.* **41**, 3913–3915 (2002).
- DeLano, W. L., *The PyMOL Molecular Graphics System* (<http://www.pymol.org/>) (2003).

Acknowledgements We thank M. Dorward for technical assistance and G. Leonard for help with data collection. J.H.N. is a Biotechnology and Biological Sciences Research Council (BBSRC) Career Development Fellow; H.D. was supported by the BBSRC; D.O.H. thanks the BBSRC for financial support; J.H.N. thanks the Wellcome Trust.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.H.N. (naismith@st-andrews.ac.uk). The nucleotide sequence of *fIA* reported here is deposited in the DDBJ, EMBL and GenBank nucleotide sequence databases under accession number AJ581748.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

US Head Office, New York

345 Park Avenue South, 10th Floor,

New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Classified Sales Representatives:

Phone: +1 800 989 7718

East USA/ Latin America/

Corporate Advertising:

Michael McGillion

Midwest USA/ Physics:

Olivia Hjalmarsson

East USA/ Canada:

Andrew Donovan

San Francisco Office

Classified Sales Representative:

Michaela Bjorkman

West USA/ West Corp. Canada

225 Bush Street, Suite 1453

San Francisco, CA 94104

Tel +1 415 781 3803

Fax +1 415 781 3805

e-mail: m.bjorkman@nature.sf.com

European Head Office, London

The Macmillan Building, 4 Crinan Street,

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayouni (4978)

Production Manager: Billie Franklin

To send materials use London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Netherlands/

Italy/ Belgium:

Patrick Phelan, Odo Wulffen

e-mail: p.phelan@nature.com

o.wulffen@nature.com

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakatsunomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: r.asami@nature.jp.com

naturejobs

Cash incentives

Britain's Institute of Physics last month announced that it intends to pay impoverished undergraduates to study physics. By doing so, it hopes to reverse two UK trends. The number of British students taking physics is falling, while student debt in the country is growing — it has risen by 43% since 2000, according to market-research company MORI. David Wallace, the institute's president, sees physics as a core discipline that is integral to Britain's knowledge economy. He hopes that providing students with an annual grant of £1,000 (US\$1,800) during their undergraduate course might encourage more to pursue the discipline.

But could this scheme not be extended? Physics isn't the only discipline facing a student shortfall in the Western world, and Britain isn't the only country where students are becoming increasingly burdened by loans. In the past few years, mathematics and chemistry have also seen a decline in student numbers — not just in Britain, but in the United States and much of mainland Europe.

Sam Rankin, director of the American Mathematical Society's Washington office, says that he is unsure whether money alone is the answer. Students care about debt, he says, but most weigh the cost of that debt against the benefit of the career it buys. They want to know what kind of job they will get after spending all those years studying, Rankin notes.

Providing cash incentives for students to move into neglected disciplines can make sense. But gathering career data and disseminating them to prospective undergraduate and graduate students will help them to choose whether their course will be a good career investment. Of course, such information could be dangerous — discouraging data might prove to be the equivalent of paying a student £1,000 not to pursue a particular discipline.

Paul Smaglik

Naturejobs editor



Contents

CAREERS AND RECRUITMENT

Opportunities in systems biology p568

CAREER VIEW

Nuts & Bolts

Selling your skills

Graduate Journal

Conference survival

Movers

Marc Tessier-Lavigne p570

WWW.NATUREJOBS.COM

Career centre

Information on the scientific job market

FOCUS

SPOTLIGHT

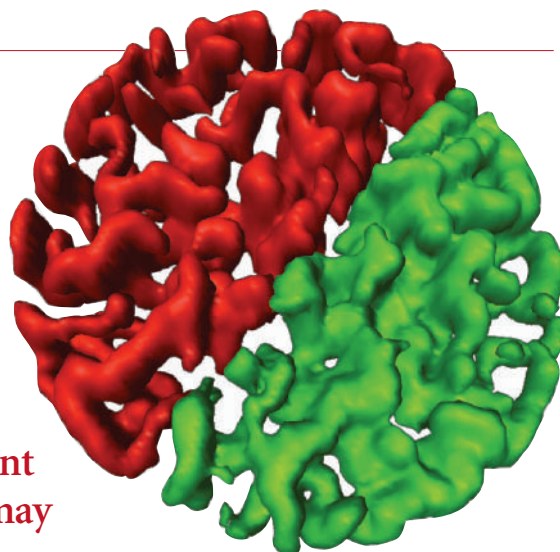
RECRUITMENT

ANNOUNCEMENTS

EVENTS

All systems go

If you can reassemble an organism from its component parts then a wealth of jobs may await you, says Hannah Hoag.



Many biologists were lured into research by the promise that, one day, they would understand how an organism worked. Over the years, they have broken cells down into their constituent parts, identifying genes and proteins and their interactions. Systems biology aims to put the organism back together again from these newly identified parts, and to predict its behaviour in nearly any situation.

Once restricted primarily to the United States, the job market for systems biology is now growing in Europe and Japan. Spacious new institutes, increased funding and a glut of positions for both experimentalists and theoreticians are providing a boost for the field — although mathematicians, physicists and engineers are receiving a warmer welcome than biologists.

Investigators commonly assume that scientists with a quantitative background will fare better in systems biology. The data deluge from fields such as genomics and proteomics is best handled by those who can work with numbers in meaningful ways. In fact, a multidisciplinary background is more appropriate. “It is not much use to hire the perfect mathematical modeller who has never heard of DNA or cells,” says Roland Eils, who heads a theoretical-bioinformatics group at the German Cancer Research Center in Heidelberg.

Biologists eager to join the field will need experience in mathematics or physics, but the ability to communicate may be most important. “A systems biologist has to be able to understand two languages: the exact language of mathematics and the fuzzy language of biology, which is more descriptive and cryptic,” says Eils.

Industry spin-offs exist in systems biology, but they remain small and tend to hire only a few researchers each year. The greatest concentration of jobs is in universities, or in institutes that straddle the line between academia and industry. In 2003, Harvard

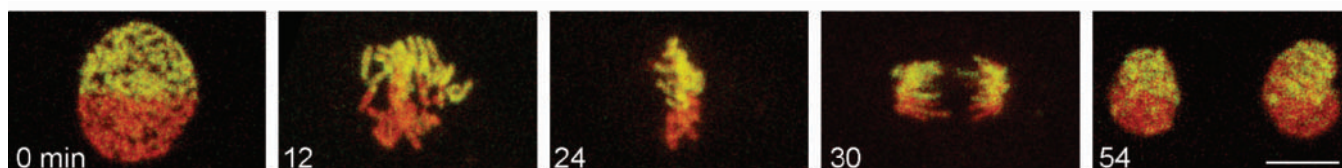
Medical School launched a systems-biology department, which will recruit more than 20 full-time faculty members. The school is the only one in the United States with a department-level systems-biology programme.

Elsewhere, multidisciplinary biocentres, both real and virtual, are the norm. Leroy Hood pioneered the movement in 2000 when he established the Institute for Systems Biology in Seattle, Washington (see *Nature* **407**, 828–829; 2000). He was the first to sever investigators’ traditional ties to a university, a move he believed was essential to foster interdisciplinary relationships.

THE PRICE OF FREEDOM

Although researchers may benefit from such autonomy, most consider project funding to be inadequate. Many of the larger grants have allowed new institutes to be built, but few grants address the large-scale nature of systems-biology projects. “Right now, funding is a challenge,” says Steve Wiley, director of the Biomolecular Systems Initiative at Pacific Northwest National Laboratory in Richland, Washington, who says that he has stitched together funding from the US Department of Energy, the National Center for Research Resources and numerous little grants to “fill in the gaps”. This patchwork of government grants, trusts and private donations can make it tough for an institute to launch a robust systems-biology programme.

But scientists remain optimistic. In the United States, the National Institutes of Health (NIH), the National Science Foundation (NSF) and the energy department are increasing their funding of systems biology. Under the NIH’s long-term funding plan, for example, \$7.4 million will be spent on initiatives, including two to four new centres. The NSF has earmarked \$4 million for quantitative systems-biology research in 2004 for the first of three years.



Nuts and bolts: these microscopy images showing how chromosomal material is split during cell division have been used to generate a mathematical model of the process.

Wiley, however, is concerned that systems-biology programmes are growing faster than the funding agencies' award structures. There is no way that a systems-biology programme could be funded by a single, investigator-initiated grant, he says.

The Japanese government, meanwhile, has provided funding for the Kitano Symbiotic Systems Project, and the recently launched Solution Oriented Research for Science and Technology Project. Hiroaki Kitano, who founded both groups, says that they are approaching systems biology from both the silicon and the cellular sides. They are working on two computer tools — systems biology mark-up language (SBML) and systems biology workbench (SBW) — that should make research more efficient, says Kitano. He hopes to hire a few experimental researchers this year.

Europe's foray into systems biology has long lagged

established later this year. In 2007, the groups will move to the Bioquant research network centre, a newly proposed institute to be affiliated with the University of Heidelberg.

SAFE HAVEN

In Britain, the Manchester Interdisciplinary Biocentre is scheduled to open in 2005. The institute will be the first large-scale centre of its kind in the country. The £35-million (US\$64-million) initiative has had large contributions from the charitable organizations the Wellcome Trust and the Wolfson Foundation. More than 500 staff, including up to 85 investigators, will pool their talents to study systems biology from the nano to the macro levels (see *Nature* **425**, 430–433; 2003).

John McCarthy, the academic coordinator of the

biocentre, sees its role as providing a haven for researchers. "A lot of people with a quantitative background don't want to move into a traditional mainstream biology department because they won't have the right environment to develop what they are working on," he says.

Finland and the Netherlands are also entering the fray. The Academy of Finland and the country's national technology agency, Tekes, have jointly funded SYSBIO, a €10-million, four-year systems-biology and bioinformatics programme that will include about 50 positions, mainly for PhD and postdoc students. Dutch universities, including the Free University of Amsterdam, are beginning to collaborate with one another

and with multinational companies, such as Unilever, to develop drugs, to understand diabetes and sleeping sickness — and to brew a better beer.

The flurry of job openings provides people interested in systems biology with international opportunities. "I am absolutely certain that we will rely on international applications," says Eils. "We don't have enough people here in Germany to fill these positions." Whereas the popularity of molecular-biology programmes has blossomed in the United States, it is the opposite in Europe, where graduates are more likely to have degrees in physiology and cell biology. Theoretical mathematicians from Eastern Europe, where molecular biology remains strong, may find themselves being wooed by US institutions.

But Eils suspects that interest will shift again and the tide will turn in a few years. "If we wait for another five years, we will have more people, especially in the field of biology, with additional training in the theoretical side," he says. "It's already happening."

Hannah Hoag is a freelance science journalist based in Montreal, Canada.



Keyed in: researchers at the German Cancer Research Center discuss a computer model of cell processes.

behind the United States and Japan, but European researchers are at last receiving some government support. Germany is doing most: both national and regional governments have recently unveiled systems-biology programmes. The education ministry will put €20 million (US\$25 million) into a national programme dedicated to the liver cell. "We decided to go with one model system that will have the most relevance for the pharmaceutical and biotech industries," says Eils. With an anticipated 78 federally funded positions to be filled by the end of the year — many at postdoc or faculty level — Germany may become the destination of choice for international researchers.

The German regional programme, boosted by the private support from the Klaus Tschira Foundation, will distribute €1.5 million annually among five centres, including the University of Heidelberg, the European Molecular Biology Laboratory, also based in Heidelberg, and the German Cancer Research Center, each of which will see a new research group

Web links

MIT Computational and Systems Biology Initiative

♦ csbi.mit.edu

Bio-X, Stanford University

♦ biox.stanford.edu/index.html

Institute for Systems Biology, Seattle

♦ www.systemsbiology.org

Pacific Northwest National Laboratory

♦ www.sysbio.org

Department of Systems Biology, Harvard Medical School

♦ sysbio.med.harvard.edu

Lewis-Sigler Institute for Integrative Genomics, Princeton University

♦ www.genomics.princeton.edu

Molecular Sciences Institute, Berkeley

♦ www.molsci.org

IBM Research

♦ www.research.ibm.com/FunGen

Manchester Interdisciplinary Biocentre

♦ www.mib.umist.ac.uk/Overview

Overview

Free University of Amsterdam

♦ www.vu.nl/home/index.cfm

German Cancer Research Center, Heidelberg

♦ www.dkfz.de/ibios/index.jsp

University of Stuttgart Systems Biology Group

♦ www.sysbio.de

Systems Biology and

Bioinformatics, Academy of Finland

♦ www.aka.fi/sysbio

Lilly Systems Biology, Singapore

♦ www.lsb.lilly.com.sg

Kitano Symbiotic Systems Project

♦ www.symbio.jst.go.jp/symbio/index.htm

GRADUATE JOURNAL

Conference survival

At the Materials Research Society meeting in Boston last October, I had flashbacks of my first conference eight years ago. In 1996, as a new master's student, I went to the Conference of Metallurgists in Montreal, Canada, and was completely overwhelmed. I didn't know what to say or who to talk to — I just had the vague sense that somehow I was supposed to 'network'. Instead, I hid from the crowds at the conference bookstore.

My refuge proved to be fortuitous. After browsing a bit, I realized the conference staff were still setting up the store. I offered to give them a hand. In return, I found an informal way to mix with scientists. The staff told me that I was not the only delegate who would rather read than talk to scientists they didn't know.

Since then, I have volunteered for a conference organizing committee and have led some student chapter meetings at conferences. And I continue to ask senior attendees how they arrived at their work, and what they like about it. It gets easier to approach new people with every conference. By helping out and talking to a few people casually, I learn more from conferences than just what is delivered at the podium. And I no longer try to hide in the bookstore. ■

Sidney Omelon is a PhD student in bone biomaterials at the Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.

NUTS & BOLTS

Selling your skills

Some countries draw a distinction between a CV (curriculum vitae) and a résumé, others deploy the descriptions interchangeably. But whether or not you view a CV as a complete summary of your scientific career and a résumé as a shorter, more targeted job application, it is clear that when seeking a new post you need to use the appropriate approach.

If you're applying for an academic position or a fellowship, stick with a traditional chronological format. Start with your name and contact information, then list your education, honours, publications, presentations, work history, research interests, professional affiliations and references with contact information. Aim for completeness — include all relevant work experience such as teaching, research, residencies and fellowships. Don't worry



With Deb Koen
Careers consultant

about length; this document can run as long as you need.

By contrast, about the only thing an application for an industrial position has in common with such a document is that it should also start with your name and contact information. Instead of focusing on education and publications, emphasize your practical experience, and tailor your experiences to the employer's needs. For example, if you are applying for a job at a drug-discovery firm that focuses on protein structures and you're from academia, say how many structures you have helped

to solve. Education is the next most important category, but you may not want to have separate entries for each rung you've climbed at the same institution. You need to keep it short — 1 to 2 pages. Attaching a separate list of publications, or providing one if asked, helps.

Regardless of the format, strive to create a document that is organized, easy to read and pleasing to the eye. Most importantly, get started now and keep it up-to-date for whenever the need arises.

For more information and samples of résumés and CVs see *Put Your Science to Work* by Peter Fiske, *Cyberspace Resume Kit* by Mary Nemnich and Fred Jandt, and www.hsph.harvard.edu (put 'sample curriculum vitae' in the search engine). ■
Deb Koen is vice-president of Career Development Services and a columnist for The Wall Street Journal's CareerJournal.com.

MOVERS

Marc Tessier-Lavigne, senior vice-president, Genentech, South San Francisco



For Marc Tessier-Lavigne, a bout of seasickness stirred up a career epiphany 24 years ago that still reverberates today. In 1980, the maths and physics major at McGill University in Montreal won a Rhodes scholarship. He planned to follow that same research direction in Oxford, UK, and he registered for a PhD in physics and selected an adviser. But while crossing

the Atlantic aboard the *QEII*, the decision — and the waves — left him feeling queasy. He realized that his interests were more varied — one of his favourite undergraduate experiences had been scraping the surface of biology while writing a thesis on biophysics — and that he still had plenty of time to learn.

"Do I really want to be on a set career path when I'm 23?" he asked himself between the Dramamines. He answered by ditching the doctorate and pursuing a bachelor's degree in a hybrid subject that no one had picked as a major for five years. "Philosophy and physiology — the connection is tenuous, but it was exactly what I wanted," he says.

That flexibility and risk-taking has since characterized his career. He declined a rare tenure-track position at University College London after he finished his doctorate there, to pursue a postdoc instead; published only review articles for three years; went after one of

the more intractable problems of neurodevelopmental biology at the time; and, last year, left a lofty academic job to jump across to industry and join biotech firm Genentech in South San Francisco (see *Nature Medicine* 10, 10; 2004).

In following such a varied path, Tessier-Lavigne unwittingly adhered to the template that is now considered the perfect background for a multidisciplinary scientist (see *Nature* 425, 542–543; 2003). He has since advised his graduate students to do a postdoc in something completely different from their thesis topic. For instance, if a student's thesis was on the cell cycle of yeast, he'd recommend against them using a fellowship to continue studying that issue — even in another organism.

Following that logic, Tessier-Lavigne is prepared to pursue research subjects he's never investigated in a setting he has never worked in. ■

CV

2001–2003: Endowed chairholder and professor, Stanford University

1991–2001: University of California, San Francisco, beginning as assistant professor of anatomy and rising to professor of anatomy and of biochemistry and biophysics by 1997

1994–2003: Howard Hughes Medical Institute, assistant investigator 1994–97, investigator 1997–2003

1987–1991: Columbia University, postdoctoral research, Center for Neurobiology and Behaviour

1987: University College London, postdoctoral researcher